ISSN 1927-7032 (Print)
ISSN 1927-7040 (Online)

International Journal of Statistics and Probability

Vol. 4, No. 4 November 2015

CANADIAN CENTER OF SCIENCE AND EDUCATION

Editorial Board

Contents

International Journal of Statistics and Probability; Vol. 4, No. 4; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Optimal Partially Replicated Cube, Star and Center Runs in Face-centered Central Composite Designs

M. P. Iwundu

Correspondence: M. P. Iwundu, Department of Mathematics and Statistics, Faculty of Physical Science and Information Technology, University of Port Harcourt, Nigeria. E-mail: mary_iwundu@yahoo.com

Received: July 30, 2015 Accepted: August 17, 2015 Online Published: September 23, 2015

doi:10.5539/ijsp.v4n4p1 URL: http://dx.doi.org/10.5539/ijsp.v4n4p1

Abstract

The variations of the Face-centered Central Composite Design under partial replications of design points are studied. The experimental conditions include replicating the cube points while the star points and center point are held fixed or not replicated, replicating the star points while the cube points and the center point are held fixed or not replicated and replicating the center point while the cube points and the star points are held fixed or not replicated. As a measure of goodness of the designs, D- and G-efficiency criteria are utilized. Results show that for the two- and three-variable quadratic models considered, the Face-centered Central Composite Design comprising of two cube portions, one star portion and a center point performed better than other variations under D-optimality criterion as well as G-optimality criterion. When compared with the traditional method of replicating the center point, the two cube portions, one star portion and a center point variation was relatively better in terms of design efficiency.

Keywords: Replication, cube points, star points, center point, D-efficiency, G-efficiency

1. Introduction

Unreplicated designs are very widely used in experimental situations. However, fitting full model for unreplicated designs results in zero degrees of freedom for error and hence tests about main and interaction effects of factors cannot be carried out. This constitutes a potential problem in statistical testing (Farrukh, 2014). Two common approaches to this problem require either pooling high-order interactions, assumed to be negligible, to estimate the error or replicating one or more experimental runs. Generally, replication of design points offers an independent and more precise estimate of experimental error.

In model building, designs with factors that are set at two levels implicitly assume that the effect of the factors on the response variable is linear and one would usually anticipate fitting the first-order model. When it is suspected that the relationship between the factors in the design and the response variable is not linear, there is the need to include one or more experimental runs. The first-order models with the presence of interaction terms are capable of representing some curvature in the response function. However, in some cases, the curvature in the response function is not adequately modeled and therefore the need to consider the second-order model for better representation.

Central Composite Designs (CCDs) originally proposed by Box & Wilson (1951) have been the practically used designs for estimating second-order response surfaces. They are so advantageous in Response Surface Methodology (RSM) for building models of the response variables without needing to carry out complete three-level factorial experiments. Applications of Central Composite Designs can be seen in various fields of study including biological, chemical, pharmaceutical fields. The CCD is particularly useful in the determination of optimum values of influential parameters of a response variable (see e.g. Alalayah *et al.* (2010)). A review of some aspects of Central Composite Designs in spherical region is presented in Chigbu *et al.* (2009).

A CCD consists of three distinct sets of experimental runs:

i. A set of factorial or fractional factorial design (cube portion) in the factors studied and each having two levels;

ii. A set of axial points (star portion);

iii. A set of center points.

In augmenting Central Composite Designs, the common practice has been the replication of only the center point

for estimation of the experimental error, improvement of the precision of the experiments and to maintain minimum number of design runs which an experimenter can afford. Two ways of replicating design are the DESIGNREP procedure which involves replicating the entire design and the POINTREP procedure which involves replicating each point in the design. When it is not possible to replicate the full design, the experimenter can obtain an estimate of pure error by replicating only some of the points in the design. One challenge of partial replications of design points is that the experimenter faces the problem of choosing the points to be replicated and the points not to be replicated in the design. Authors including Cochran and Cox (1957), Montgomery (1997) and Atkinson and Donev (1992) have discussed extensively the analysis of such replicated experiments.

Quite recently, many experimenters have focused on the effect of replicating the non-center points as against the usual replication of the center point in exploring response surfaces. Chigbu and Ohaegbulem (2011) considered the preference of replicating factorial runs to axial runs in restricted second-order designs. They observed in general that under orthogonality and rotatability restrictions, the replicated cubes plus one star variation was better than the one cube plus replicated star variation in the sense of D-optimality. The number of experimental runs employed was $N = n_1 2^k + n_2 2k + n_0$ where n_1 is the number of cubes, n_2 is the number of stars and n_0 is the number of center point. Although allowing for partial replication of the cubes and the stars, every point in the cube as well as the star was utilized. Ukaegbu and Chigbu (2014) considered the performance of the partially replicated cube and star portions of orthogonal Central Composite Designs in spherical regions. One particular focus was the replication of the cube and star portions without replicating the center point in k-factor experiment. Also, the performance of the Central Composite Designs with respect to stability, small predictive variance and prediction capability was studied using graphical techniques and single-value optimality criteria. Results indicate that replicating the star portions of the Central Composite Designs considerably reduces the prediction variance and thus improves G-efficiency than replicating the cube portion.

Oyejola and Nwanya (2015) considered the performance of five varieties of Central Composite Design when the axial portions are replicated and the center point increased one and three times. Ahn (2015) devised a new CCD called the CCD-R for experiments not just at the center but also at non-center points. The flexibility of the CCD-R is seen in the existence of a myriad of perfectly orthogonal and nearly rotatable designs. Ahn (2015) considered that when a two-level full or fractional experiment is conducted, a few center runs would be adequate to detect the quadratic effects over the region of exploration. However, in situations where the parameters of quadratic model are to be separately estimated, more runs at some more design points are needed. In addressing this problem, the augmentation of the two-level full or fractional factorial design with a center and 2k axial points was proposed, where k is the number of independent factors in the experiment.

In this work, the effect of partially replicating the factorial points and the star points of the Face-centered Central Composite Designs with respect to replicating the center point on response surface designs is investigated. This requires

i. Constructing partially replicated exact designs for two and three variable quadratic models.

ii. Assessing the goodness of the designs using two single-value criteria, namely D- and G-efficiency criteria.

For two input variables (i.e. k = 2), the Face-centered Central Composite Design consists of n_c center points, four factorial points and four axial points. For three input variables (i.e. k = 3), the Face-centered Central Composite Design consists of n_c center points, eight factorial points and six axial points. The axial points are parallel to each variable axis on a circle of radius, $\alpha = 1.0$ and origin at the center point. The designated α is the radius which determines the geometry and defines a square for two input variables and a cube for three input variables. According to Montgomery (1997) and Zahran (2002). Face-centered Central Composite Design is the most useful cuboidal region in practice because it requires only three levels of each factor.

Draper and Guttman (1988) observed that the adequacy of an experimental design can be determined from the information matrix. Some criteria that are based on the information matrix include A-, D-, E-, G- and I-optimality criteria. Rady *et al.* (2009) gave a concise survey on the optimality criteria with particular attention on relationships among the several optimality criteria. Following the definitions of Atkinson and Donev (1992), A-optimality criterion seeks to minimize the trace of the variance-covariance matrix. This criterion results in minimizing the average variance of the estimated regression coefficients. D-optimality criterion maximizes the amount of information in an experimental design. As assessed by the information matrix, D-optimality criterion maximizes the determinant of information matrix of the design and equivalently minimizes the determinant of the variance-covariance matrix. Hence for a specified model, a D-optimal design minimizes the variances of parameter estimates as well as the covariances between parameter estimates. On the other hand, G-optimality criterion minimizes the maximium variance of prediction over the design space.

A number of standard measures have been proposed in the literature to summarize the efficiency of a design. Some of these measures can be seen in Atkinson and Donev (1992), Wong (1994) and Chukwu and Yakubu (2012). To assess the goodness of designs, two single-value efficiency criteria, namely, the D- and G-efficiencies are commonly employed. As in the literature on optimal designs, efficiency values lie between zero and one, a design having efficiency value of 1.0 implies that the design is 100% efficient. Hence in comparing designs, a design with a higher efficiency value would be preferred. According to Atkinson and Donev (1992), D-efficiency of an arbitrary design, ξ_N, over an optimal design, ξ_N^* is defined as

$$D_{eff} = \frac{M(\xi_N)}{M(\xi_N^*)} .$$

The G-efficiency of an arbitrary design, ξ_N, is defined as

$$G_{eff} = \frac{d(\xi_N^*)}{d(\xi_N)} = \frac{p}{d(\xi_N)} ;$$

Where $d(\xi_N^*)$ is the maximum variance of predicted response associated with ξ_N^* and $d(\xi_N)$ is the maximum variance of predicted response associated with ξ_N.

Here, p is the number of model parameters and N is the number of requested runs. The D-efficiency can be interpreted as the relative number of runs (in percent) that would be required by an orthogonal design to achieve the same value of the determinant $|X^TX|$. In practice, an orthogonal design may not be possible in many cases; hence orthogonality becomes only a theoretical "yard-stick." Therefore, one should use D-efficiency measure rather as a relative indicator of efficiency to compare other designs. D-efficiency measure relates to D-optimality criterion as G-efficiency measure relates to the G-optimality criterion, which concentrates on minimizing the maximum value of the standard error of the predicted response.

2. Method

In this work, the variation of the Central Composite Design (CCD) is studied when

(i) The cube points are replicated while the star points and center point are held fixed or not replicated;

(ii) The star points are replicated while the cube points and the center point are held fixed or not replicated;

(iii) The center point is replicated while the cube points and the star points are held fixed or not replicated.

Efficiencies of the constructed designs are assessed using D- and G- efficiency criteria.

In studying the partial replication of Central Composite Design, the second-order polynomial model in equation (1) is employed.

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j>i}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_{ii}^2 + \varepsilon \quad (1)$$

This model can be rewritten as

$$Y = X\beta + \varepsilon \quad (2)$$

Where

Y is the Nx1 vector of observed values

X is the design matrix

β is the px1 vector of unknown parameters which are estimated on the basis of N uncorrelated observations.

ε is the random additive error associated with Y and is independently and identically distributed with zero mean and constant variance.

To explore the Face-centered Central Composite Design with partial replication of the cube or the factorial points, we observe that the k-variable second-order full model has p model parameters given by

$$p = \frac{(k+1)(k+2)}{2} \quad (3)$$

The factorial portion of the Central Composite Design comprises of experimental runs of the 2^k factorial design.

For $k = 2$, the experimental runs are

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

For k = 3, the experimental runs are

$$V = \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & +1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ -1 & +1 & +1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \end{pmatrix}$$

The star portion of the Central Composite Design comprises of the experimental runs

$$S = \begin{pmatrix} \alpha & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ -\alpha & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \alpha & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & -\alpha & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cdots & \alpha \\ 0 & 0 & 0 & 0 & \cdots & \cdots & -\alpha \end{pmatrix}$$

For k = 2, this becomes

$$S = \begin{pmatrix} \alpha & 0 \\ -\alpha & 0 \\ 0 & \alpha \\ 0 & -\alpha \end{pmatrix}$$

For k = 3,

$$S = \begin{pmatrix} \alpha & 0 & 0 \\ -\alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & -\alpha & 0 \\ 0 & 0 & \alpha \\ 0 & 0 & -\alpha \end{pmatrix}$$

The center portion of the Central Composite Design comprises of the experimental run

$$C = (0 \quad 0 \quad \ldots \quad 0).$$

For k = 2, this becomes

$$C = (0 \quad 0)$$

For k = 3, it becomes

$$C = (0 \quad 0 \quad 0).$$

The information matrix of a CCD shall be expressed in terms of the number of cube points, star points and center point. Thus, the number of experimental runs is given by $N = n_1 2^k + \binom{n11}{r} + n_2 2k + \binom{n22}{r} + n_0$ where n_1 is the number of cube portions, n_2 is the number of star portions, n_0 is the number of center points, n_{11} refers to the number of cube points in the cube portion of the CCD and n_{22} refers to the number of star points in the star portion of the CCD. For the purpose of this work n_1 and n_2 are set at unity, $V+\binom{n11}{r}$ implies taking the cube portion and additional r distinct cube points from the available n_{11} cube points, $S+\binom{n22}{r}$ implies taking the star portion and additional r distinct star points from the available n_{22} star points and C+2 implies taking the center point and additional two center points.

For k = 2, the various variations or experimental conditions to study in replicating the vertex points while the star points and center point are held fixed or not replicated are as tabulated in Table 1.

Table 1. Variations for replicating the vertex points (k = 2)

Experimental Condition			Design Size N
Vertex	**Star**	**Center**	
$V+^4C_4$	S	C	13
$V+^4C_3$	S	C	12
$V+^4C_2$	S	C	11
$V+^4C_1$	S	C	10

In replicating the star points while the vertex points and center point are held fixed or not replicated, the various variations or experimental conditions to study are as tabulated in Table 2.

Table 2. Variations for replicating the star points (k = 2)

Experimental Condition			Design Size N
Vertex	**Star**	**Center**	
V	$S+^4C_4$	C	13
V	$S+^4C_3$	C	12
V	$S+^4C_2$	C	11
V	$S+^4C_1$	C	10

In replicating the center point while the vertex points and star points are held fixed or not replicated, the various variations or experimental conditions to study are as tabulated in Table 3

Table 3. Variations for replicating the center point (k = 2)

Experimental Condition			Design Size N
Vertex	**Star**	**Center**	
V	S	C+4	13
V	S	C+3	12
V	S	C+2	11
V	S	C+1	10

For k = 3, the various variations or experimental conditions to study in replicating the vertex points while the star points and center point are held fixed or not replicated are as tabulated in Table 4.

Table 4. Variations for replicating the vertex points (k = 3)

Experimental Condition			Design Size N
Vertex	**Star**	**Center**	
$V+^8C_8$	S	C	23
$V+^8C_7$	S	C	22
$V+^8C_6$	S	C	21
$V+^8C_5$	S	C	20
$V+^8C_4$	S	C	19
$V+^8C_3$	S	C	18
$V+^8C_2$	S	C	17
$V+^8C_1$	S	C	16

In replicating the star points while the vertex points and center point are held fixed or not replicated, the various variations or experimental conditions to study are as tabulated in Table 5

Table 5. Variations for replicating the star points (k = 3)

Experimental Condition			Design Size N
Vertex	**Star**	**Center**	
V	$S+^8C_8$	C	23
V	$S+^8C_7$	C	22
V	$S+^8C_6$	C	21
V	$S+^8C_5$	C	20
V	$S+^8C_4$	C	19
V	$S+^8C_3$	C	18
V	$S+^8C_2$	C	17
V	$S+^8C_1$	C	16

In replicating the center point while the vertex points and star points are held fixed or not replicated, the various variations or experimental conditions to study are as tabulated in Table 6

Table 6. Variations for replicating the star points (k = 3)

Experimental Condition			Design Size N
Vertex	**Star**	**Center**	
V	S	C+8	23
V	S	C+7	22
V	S	C+6	21
V	S	C+5	20
V	S	C+4	19
V	S	C+3	18
V	S	C+2	17
V	S	C+1	16

For each experimental condition, an *N*-point design shall be chosen to maximize the determinant of information matrix. Onukogu and Iwundu (2007), Madukaife and Oladugba (2010) and Iwundu and Albert-Udochukwuka (2014) have provided helpful rules for selecting design points to maximize the determinant of information matrix.

Let

$$\xi_N = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{pmatrix}$$

be an *N*-point design measure depending on k-variable quadratic model, having p-parameters. The *N*xp design matrix

$$\mathrm{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nk} \end{pmatrix}$$

gives the values of independent variables that are used in the statistical models and further contains the column of 1's that represent the intercept term as well as the columns for the products and powers associated with other model terms. The pxp information matrix, M, associated with ξ_N is obtained from $\mathrm{X^TX}$ and normalized as $\frac{1}{N}\mathrm{X^TX}$, where the notation, $(.)^{\mathrm{T}}$ represents transpose. The criterion that allows maximization of determinant of information matix of a design is the D-optimality criterion.

Let ξ_N^1, ξ_N^2, ... , ξ_N^m be m design measures defined on the design region of the Face-centered Central Composite Design and having non-singular information matrices M_1, M_2, ... , M_m, respectively. The design measure ξ_N^1 is preferred, in terms of D-optimality criterion, to the design measuresξ_N^2, ..., ξ_N^m iff the determinant

$$\mathrm{Det}\ (M_1) = \max\ \{\mathrm{Det}\ (M_1), \mathrm{Det}\ (M_2), \ldots , \mathrm{Det}\ (M_m)\}.$$

Also let $\underline{x}_i = (1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ik})$; i = 1, 2, ... , N be the i[th] row of the design matrix X, associated with the design point $(x_{i1} \quad x_{i2} \quad \cdots \quad x_{ik})$. The variance of prediction, $V\{y(\underline{x}_i)\}$, at the i[th] design point $\underline{x}_i = (1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ik})$ is

$$V\{y(\underline{x}_i)\} = (1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ik})\ M^{-1}(1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ik})^T$$

The criterion that allows minimization of the maximum predictive variance is the G-optimality criterion.

Suppose

$V^1 = V\{y(\underline{x}_1)\}$ is the maximum variance of prediction associated with the design measure ξ_N^1,

$V^2 = V\{y(\underline{x}_2)\}$ is the maximum variance of prediction associated with the design measure ξ_N^2,

⋮

$V^m = V\{y(\underline{x}_m)\}$ is the maximum variance of prediction associated with the design measure ξ_N^m.

The design measure ξ_N^1 is preferred in terms of G-optimality criterion to the design measures ξ_N^2, ... , ξ_N^m iff

$$V\{y(\underline{x}_1)\} = \min \{ V\{y(\underline{x}_2)\}, V\{y(\underline{x}_2)\}, \ldots , V\{y(\underline{x}_m)\}\}.$$

3. Results

Using the second-order polynomial model in equation (1), the partial replications of the factorial points and the star points with respect to replicating the center point are investigated with the following results.

3.1 Two-Factor Partially Replicated Central Composite Design

In exploring the two-factor Face-centered Central Composite Design with partial replication of the cube or factorial points, it is observed that the two-variable second-order full polynomial model has six model parameters. For the Face-centered Central Composite Design in two variables, there are basically nine design points or experimental runs. The cube points otherwise called vertex or factorial points

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

are denoted V.

The axial or star points

$$\begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}$$

are denoted S.

The center point, $(0 \quad 0)$, is denoted C.

Case I: Replicating the vertex points while the star points and center point are held fixed or not replicated.

Using the experimental conditions in Table 1, partially replicated exact designs of size N = 13, 12, 11, 10 are constructed.

The design measure for N = 13 is

$$\xi_{13} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 1 & -1 \\ -1 & -1 \\ 1 & 1 \\ -1 & 1 \\ 1 & -1 \\ -1 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}$$

For the six parameter model, the design matrix is

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The corresponding information matrix is

$$M = \frac{1}{N} X^T X = \begin{pmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.7692 & 0.7692 \\ 0.0000 & 0.7692 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.7692 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.7692 & 0.0000 & 0.0000 \\ 0.7692 & 0.0000 & 0.0000 & 0.0000 & 0.7692 & 0.7692 \\ 0.7692 & 0.0000 & 0.0000 & 0.0000 & 0.7692 & 0.7692 \end{pmatrix}$$

The determinant value of the information matrix is

$$\text{Det M} = 0.01127$$

The variance of prediction at each design point of ξ_{13} is, respectively

$$V_1 = 5.7544$$
$$V_2 = 5.7544$$
$$V_3 = 5.7544$$
$$V_4 = 5.7544$$
$$V_5 = 5.7544$$
$$V_6 = 5.7544$$
$$V_7 = 5.7544$$
$$V_8 = 5.7544$$
$$V_9 = 8.1824$$
$$V_{10} = 6.2706$$
$$V_{11} = 6.2706$$
$$V_{12} = 6.2706$$
$$V_{13} = 6.8824$$

The maximum predictive variance is 8.1824.

The design measure for $N = 12$ is

$$\xi_{12} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 1 & -1 \\ -1 & -1 \\ -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}$$

For the six parameter model, the design matrix is

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The associated information matrix is

$$M = \frac{1}{N} X^T X = \begin{pmatrix} 1.0000 & -0.083 & -0.083 & -0.083 & 0.7500 & 0.7500 \\ -0.083 & 0.7500 & -0.083 & -0.083 & -0.083 & -0.083 \\ -0.083 & -0.083 & 0.7500 & -0.083 & -0.083 & -0.083 \\ -0.083 & -0.083 & -0.083 & 0.5830 & -0.083 & -0.083 \\ 0.7500 & -0.083 & -0.083 & -0.083 & 0.7500 & 0.5830 \\ 0.7500 & -0.083 & -0.083 & -0.083 & 0.5830 & 0.7500 \end{pmatrix}$$

The determinant value of the information matrix is

$$\text{Det M} = 0.0102$$

The variance of prediction at each design point is, respectively

$$V_1 = 9.5303$$
$$V_2 = 5.3129$$
$$V_3 = 5.3129$$
$$V_4 = 5.3509$$
$$V_5 = 5.3129$$
$$V_6 = 5.3509$$
$$V_7 = 5.3129$$
$$V_8 = 5.1488$$
$$V_9 = 5.8955$$
$$V_{10} = 5.1488$$
$$V_{11} = 5.8955$$
$$V_{12} = 6.4274$$

The maximum predictive variance is 9.5303.

The design measure for $N = 11$ is

$$\xi_{11} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 1 & -1 \\ -1 & -1 \\ -1 & 1 \\ 1 & 1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}$$

For the six parameter model, the design matrix is

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The associated information matrix is

$$M = \begin{pmatrix} 1.0000 & 0.0000 & 0.1818 & 0.0000 & 0.7272 & 0.7272 \\ 0.0000 & 0.7272 & 0.0000 & 0.1818 & 0.0000 & 0.0000 \\ 0.1818 & 0.0000 & 0.7272 & 0.0000 & 0.1818 & 0.1818 \\ 0.0000 & 0.1818 & 0.0000 & 0.5454 & 0.0000 & 0.0000 \\ 0.7272 & 0.0000 & 0.1818 & 0.0000 & 0.7272 & 0.5454 \\ 0.7272 & 0.0000 & 0.1818 & 0.0000 & 0.5454 & 0.7272 \end{pmatrix}$$

The determinant value of the information matrix is

$$\text{Det } M = 0.00954.$$

The variance of prediction at each design point is, respectively

$$V_1 = 4.9063$$
$$V_2 = 4.9063$$
$$V_3 = 8.7396$$
$$V_4 = 8.7396$$
$$V_5 = 4.9063$$
$$V_6 = 4.9063$$
$$V_7 = 5.5000$$
$$V_8 = 5.9583$$
$$V_9 = 5.7396$$
$$V_{10} = 5.7396$$
$$V_{11} = 5.9583$$

The maximum predictive variance is 8.7396.

For $N = 10$

$$\xi_{10} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 1 & -1 \\ -1 & -1 \\ -1 & 1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 0 & 0 \end{pmatrix}$$

For the six parameter model, the design matrix is

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The associated information matrix is

$$M = \begin{pmatrix} 1.0000 & -0.100 & 0.1000 & -0.100 & 0.7000 & 0.7000 \\ -0.100 & 0.7000 & -0.100 & 0.1000 & -0.100 & -0.100 \\ 0.1000 & 0.1000 & 0.7000 & -0.100 & 0.1000 & 0.1000 \\ -0.100 & 0.1000 & -0.100 & 0.5000 & -0.1000 & -0.1000 \\ 0.7000 & -0.100 & 0.1000 & -0.100 & 0.7000 & 0.5000 \\ 0.7000 & -0.100 & 0.1000 & -0.100 & 0.5000 & 0.7000 \end{pmatrix}$$

The determinant value of the information matrix is

$$\text{Det } M = 0.00936$$

The variance of prediction at each design point is, respectively

$$V_1 = 4.4615$$
$$V_2 = 8.0513$$
$$V_3 = 7.9487$$
$$V_4 = 8.0513$$
$$V_5 = 4.4615$$
$$V_6 = 5.2821$$
$$V_7 = 5.4872$$
$$V_8 = 5.4872$$
$$V_9 = 5.2821$$
$$V_{10} = 5.4872$$

The maximum predictive variance is 8.0513.

Case II: Replicating the star points while the vertex points and center point are held fixed or not replicated.

Using the experimental conditions in Table 2, partially replicated exact designs of size N = 13, 12, 11, 10 are constructed.

The design measures for the respective N-point exact designs are;

$$\xi_{13} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\xi_{12} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\xi_{11} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$\xi_{10} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

For case II, the associated maximum determinant values and maximum variances of prediction are as tabulated in Table 7.

Table 7. Maximum determinant value and maximum predictive variances for Case II, k = 2

Design Size N	Maximum determinant value of information matrix	Maximum variance of prediction
13	0.005940	9.2857
12	0.006344	9.0405
11	0.00705	8.67307
10	0.00806	7.9762

Case III: Replicating the center point while the vertex points and star points are held fixed or not replicated.

Using the experimental conditions in Table 3, partially replicated exact designs of size *N* = 13, 12, 11, 10 are constructed.

The design measures for the respective *N*-point exact designs are;

$$\xi_{13} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\xi_{12} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\xi_{11} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$\xi_{10} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \\ -1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

For Case III, the associated maximum determinant values and maximum variances of prediction are as tabulated in Table 8.

Table 8. Maximum determinant value and maximum predictive variances for Case III, k = 2

Design Size N	Maximum determinant value of information matrix	Maximum variance of prediction
13	0.00346	10.2730
12	0.00634	9.5000
11	0.00618	8.7325
10	0.00806	7.9762

3.2 Three-Factor Partially Replicated Central Composite Design

In exploring the three-factor partially replicated Central Composite Design, it is observed that the three-variable second-order full polynomial model has ten model parameters. For the Face-centered Central Composite Design in three variables, the eight factorial points

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$ are denoted V.

The six axial or star points

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$ are denoted S.

The center point $(0 \quad 0 \quad 0)$ is denoted C.

Case I: Replicating the vertex points while the star points and center points are held fixed or not replicated.

Using the experimental conditions in Table 4, we construct partially-replicated exact designs of size N = 23, 22, … , 16.

The design measure for $N = 23$ is

$$\xi_{23}= \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \\ -1 & -1 & -1 \\ -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

For the ten parameter model, the design matrix is

$$X = \begin{pmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The associated information matrix is

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7826 & 0.7826 & 0.7826 \\ 0 & 0.7826 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7826 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.7826 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.6956 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.6956 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.6956 & 0 & 0 & 0 \\ 0.7826 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7826 & 0.6956 & 0.6956 \\ 0.7826 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6956 & 0.7826 & 0.6956 \\ 0.7826 & 0 & 0 & 0 & 0 & 0 & 0 & 0.6956 & 0.6956 & 0.7826 \end{pmatrix}$$

The determinant of information matrix is

$$\text{Det } M = 0.0004106$$

The variance of prediction at each design point is, respectively

$V_1 = 9.5672$
$V_2 = 9.5672$
$V_3 = 9.5672$
$V_4 = 9.5672$
$V_5 = 9.5672$
$V_6 = 9.5672$
$V_7 = 9.5672$
$V_8 = 9.5672$
$V_9 = 9.5672$
$V_{10} = 9.5672$
$V_{11} = 9.5672$
$V_{12} = 9.5672$
$V_{13} = 9.5672$
$V_{14} = 9.5672$
$V_{15} = 9.5672$
$V_{16} = 9.5672$
$V_{17} = 11.7441$
$V_{18} = 11.7441$
$V_{19} = 11.7441$
$V_{20} = 11.7441$
$V_{21} = 11.7441$
$V_{22} = 11.7441$
$V_{23} = 6.4607$

The maximum variance of prediction is 11.7441

The results for $N = 22, 21, \ldots, 16$ are as tabulated in Table 9.

Table 9. Maximum determinant values and maximum predictive variances for Case I, k = 3

Design Size N	Maximum determinant value of information matrix	Maximum variance of prediction
23	0.00041	11.7441
22	0.00037	15.6689
21	0.00034	15.2767
20	0.00032	14.8206
19	0.00031	14.2856
18	0.00029	14.0184
17	0.00029	13.3526
16	0.00031	12.6714

Case II: Replicating the star points while the vertex points and center point are held fixed or not replicated.

Using the experimental conditions in Table 5, partially replicated exact designs of size N = 21, 20, ... , 16 are constructed. As with Case I, the best *N*-point exact design is obtained and the process continues. The required computations yield the results for *N* = 21, 20, ... , 16 as tabulated in Table 10.

Table 10. Maximum determinant values and maximum predictive variances for Case II, k = 3

Design Size N	Maximum determinant value of information matrix	Maximum variance of prediction
21	0.0001187	15.9089
20	0.0001465	15.1729
19	0.0001662	14.6312
18	0.0001938	14.0591
17	0.0002211	13.2672
16	0.0002608	12.6742

Case III: Replicating the center point while the vertex points and star points are held fixed or not replicated.

Using the experimental conditions in Table 6, partially replicated exact designs of size *N* = 23, 22, ... , 16 are constructed. As with Cases I and II, the best *N*-point exact design is obtained and the process continues. The required computations yield the results for *N* = 23, 22, ... , 16 as tabulated in Table 11.

Table 11. Maximum determinant values and maximum predictive variances for Case III, k = 3

Design Size N	Maximum determinant value of information matrix	Maximum variance of prediction
23	0.0000147	18.2263
22	0.0000209	17.4382
21	0.0000302	16.6506
20	0.0000440	15.8636
19	0.0000648	15.0776
18	0.0000964	14.2929
17	0.0000144	13.5102
16	0.0000216	12.7310

In assessing the goodness of the constructed optimal exact designs we compute the D-efficiency and G-efficiency values as tabulated in Tables 12 and 13 for k = 2 and k = 3, respectively.

Table 12. Optimal values and D- and G-efficiency values (k = 2)

Experimental Condition			Design Size N	Determinant of Information matrix	Maximum variance of prediction	D-efficiency	G-efficiency
Vertex point	Star point	Center point					
V+$^4C_4$	S	C	13	0.0113	6.8824	1.0000	1.0000
V+$^4C_3$	S	C	12	0.0102	9.5303	0.9831	0.7222
V+$^4C_2$	S	C	11	0.0095	8.7396	0.9715	0.7875
V+$^4C_1$	S	C	10	0.0094	8.0513	0.9698	0.8548
V	S+$^4C_4$	C	13	0.0059	9.2857	0.8974	0.7412
V	S+$^4C_3$	C	12	0.0063	9.0405	0.9072	0.6637
V	S+$^4C_2$	C	11	0.0071	8.6731	0.9255	0.7935
V	S+$^4C_1$	C	10	0.0081	7.9762	0.9460	0.8629
V	S	C+4	13	0.0035	10.2730	0.8226	0.6700
V	S	C+3	12	0.0046	9.5000	0.8609	0.7245
V	S	C+2	11	0.0062	8.7325	0.9048	0.7881
V	S	C+1	10	0.0081	7.9762	0.9460	0.8629

Table 13. Optimal values and D- and G-efficiency values (k = 3)

Experimental Condition			Design Size N	Determinant of Information matrix	Maximum variance of prediction	D-efficiency	G-efficiency
Vertex point	Star point	Center point					
$V+^{8}C_{8}$	S	C	23	0.0004106	11.7441	1.0000	1.0000
$V+^{8}C_{7}$	S	C	22	0.0003740	15.6689	0.9907	0.7495
$V+^{8}C_{6}$	S	C	21	0.0003447	15.2767	0.9827	0.7688
$V+^{8}C_{5}$	S	C	20	0.0003225	14.8206	0.9761	0.7924
$V+^{8}C_{4}$	S	C	19	0.0003075	14.2856	0.9715	0.8221
$V+^{8}C_{3}$	S	C	18	0.0002968	14.0184	0.9681	0.8378
$V+^{8}C_{2}$	S	C	17	0.0002945	13.3526	0.9673	0.8795
$V+^{8}C_{1}$	S	C	16	0.0003013	12.6714	0.9695	0.9268
V	$S+^{6}C_{6}$	C	21	0.0001187	15.9089	0.8833	0.7382
V	$S+^{6}C_{5}$	C	20	0.0001465	15.1729	0.9021	0.7740
V	$S+^{6}C_{4}$	C	19	0.0001662	14.6312	0.9135	0.8027
V	$S+^{6}C_{3}$	C	18	0.0001938	14.0591	0.9277	0.8353
V	$S+^{6}C_{2}$	C	17	0.0002211	13.3375	0.9400	0.8805
V	$S+^{6}C_{1}$	C	16	0.0002608	12.6742	0.9556	0.9266
V	S	C+8	23	0.0000147	18.2263	0.7168	0.6443
V	S	C+7	22	0.0000209	17.4382	0.7425	0.6735
V	S	C+6	21	0.0000302	16.6506	0.7703	0.7053
V	S	C+5	20	0.0000440	15.8636	0.7998	0.7403
V	S	C+4	19	0.0000648	15.0776	0.8314	0.7789
V	S	C+3	18	0.0000964	14.2929	0.8651	0.8217
V	S	C+2	17	0.0000144	13.5102	0.9005	0.8693
V	S	C+1	16	0.0000216	12.7310	0.9378	0.9225

4. Discussion

In addressing the problem of partially replicated cube, star and center runs for estimation of error degrees of freedom in Response Surface Methodology**,** emphasis should not be on the replication of only center point as the replication of non-center points performs credibly well. Design optimality plays a major role in the choice of experimental designs. As observed in the study on the effects of partially replicating the factorial points and the star points of the Face-centered Central Composite Designs with respect to replicating the center points, replicating the cube points offered better designs as measured by the D- and G-efficiency values than replicating the center point. This signifies the preference of replicating non-center points, particularly the cube points.

Specifically, for two-variable quadratic model, the D-optimal exact design was observed under the experimental condition $(V+^{4}C_{4})+S+C$, which implies the replication of cube points. This design also had the minimum maximum variance of prediction over all designs considered. In comparison with designs under the varying experimental conditions, the design comprising of two cube portions, one star portion and one center point was more efficient in terms of D- and G-efficiencies. The implication is that replicating cube points allows more precise estimate of model parameters as the variances of the model parameters are minimized and the covariances between the model parameters are minimized. Furthermore, replicating cube points allows minimization of the maximum variance of prediction over the design space.

For three-variable quadratic model, the design comprising of two cube portions, one star portion and a center point performed better than other combinations in terms of D-optimality criterion as well as G-optimality criterion. The D- and G-optimal exact designs were observed using the design comprising of two cube portions, one star portion and a center point. This again implies the preference of replicating the cube points. The design comprising of two cube portions, one star portion and a center point had the maximum determinant value of information matrix as well as the minimum maximum variance of prediction over all designs considered. Again, replicating cube points allowed a more precise estimate of model parameters as the variances of the model parameters are minimized and the covariances between the model parameters are minimized. As with the two-variable model, replicating cube points allowed minimization of the maximum variance of prediction over the design space.

For cases under study, the best D-efficiency value was associated with replicating the cube points and the best D-efficiency value was associated with replicating the cube points. In fact, the lowest D-efficiency value associated with replicating the cube points was still better than the highest D-efficiency value associated with replicating the center point. This was generally true for G-efficiency. For each quadratic model considered, the efficiencies of the designs were computed relative to the best design within a class of designs. Specifically, the best D-optimal design for two-variable quadratic model was obtained and the D-efficiencies of other designs were computed relative to this best D-optimal design. Similarly, the best G-optimal design for two-variable quadratic model was obtained and the G-efficiencies of other designs were computed relative to this best G-optimal design. As with the two-variable quadratic model, the efficiencies of the designs for the three-variable quadratic model were computed relative to the best design within a class of designs. Hence, the best D-optimal design for three-variable quadratic model was obtained and the D-efficiencies of the other designs were computed relative to this best D-optimal design. Similarly, the best G-optimal design for three-variable quadratic model was obtained and the G-efficiencies of the design were computed relative to this best G-optimal design.

Although there was no consideration on A-efficiency criterion, designs that were D- and G-efficient also maximized the trace of the information matrix thereby minimizing the trace of the variance-covariance matrix. This shows that by replicating the cube points, the average variance of parameter estimates are minimized. For two- and three-variable quadratic models considered, the design comprising of two cube portions, one star portion and a center point, that maximized the determinant of information matrix as well as minimizing the maximum variance of prediction also maximized the trace of the information matrix with trace value of 4.6922 for the two-variable model and trace value of 7.7824 for the three-variable model. In partial replication of design points, complete replication of cube portion offered better designs as measured by the efficiency values than replicating some design points of the cube portion.

5. Conclusion

The effects of partially replicating the non-center points, with respect to replicating the center point of the Face-centered Central Composite Designs were considered using two- and three-variable quadratic models. As a measure of goodness of the designs, D- and G-efficiency single-value criteria were utilized. In all cases considered, the experimental designs associated with replicating only the center point were not as efficient as replicating the cube points in terms of D- and G-efficiency. We recommend that emphasis should shift away from replication of only center points when using response surface designs in optimizing response variables, as non-center points perform credibly well. However, the concepts of rotatability and orthogonality of the designs should be imposed.

References

Ahn, H. (2015). Central Composite Design for the experiments with replicate runs at factorial and axial points. Department of Industrial Engineering, Scokyeong University, Seoul, Korea 136-704. Link:springer.com/chapter10/10.1007%2F978-3-662-47200-2_101. http://dx.doi.org/10.1007/978-3-662-47200-2_101

Alalayah, W. M., Kalil, M. S., Kadhum, A. A. H., Jahim, J., Zaharim, A., Alauj, N. M., & Elshafie, A. (2010). Applications of the Box-Wilson design model for Bio-hydrogen production using Clostridium Saccharoperbutylacetonicum N1-4 (ATCC 13564). *Pakistan Journal of Biological Sciences, 13,* 674-682. http://dx.doi.org/10.3923/pjbs.2010.674.682 URL:http://scialert.net/abstract/?doi=pjbs.2010.674.682

Atkinson, A. C., & Donev, A. N. (1992). On Optimum Experimental Designs. Oxford statistical science series, clarendon press.

Box, G. E. P., & Wilson, K. B. (1951). On the Experimental Attainment of Optimum conditions. *Journal of the Royal Statistical society. Series B*(1), 1-45.

Chigbu, P. E., & Ohaegbulem, E. U. (2011). On the Preference of Replicating Factorial Runs to Axial Runs in Restricted Second Order Design. *Journal of Applied Sciences, 11*(22), 3732-3737.

Chigbu, P. E, Ukaegbu, E. C. and Nwanya, J. C. (2009). On comparing the prediction variances of some Central Composite Designs in spherical region: A review. Statistica, anno, LXIX(4), 285-298.

Chukwu, A. U., & Yakubu, Y. (2012). Comparison of optimality criteria of reduced models for response surface designs with restricted randomization. *Progress in Applied Mathematics, 4*(2), 110-126.

Cochran, W. G., & Cox, G. M. (1957). Experimental designs, Second Edition, John Wiley & Sons, New York.

Draper, N. R., & Gutman, I. (1988). An index of rotatability, Technometrics, 30, pp. 105-111.

Farrukh, J. (2014). Searching for Optimum: 2^k Factorial Design with Added Central Points, Central Composite Designs, and Response Surface Methods. www.maths.lth.se>kurser>fms072_vt14

Montgomery, D. C. (1997). Design and analysis of Experiments, (4^{th}ed.) John Wiley & Sons, New York.

Oyejola, B. A., & Nwanya, J. C. (2015). Selecting the right Central Composite Design. *International Journal of Statistics and Applications, 5*(1), 21-30.

Rady, E. A., Abd El-Monsef, M. M. E., & Seyam, M. M. (2009). Relationship among several optimality criteria, interstat.statjournals.net>YEAR>articles.

Ukaegbu, E. C., & Chigbu, P. E. (2014). Graphical evaluation of the prediction capabilities of partially replicated orthogonal Central Composite Design. *Quality and Reliability Engineering, 31*(4). http://dx.doi.org/10.1002/qre.1630

Wong, W. K. (1994). Comparing robust properties of A, D, E and G-optimal designs. *Computational Statistics & Data Analysis, 18*, 441-448.

Zahran, A. R. (2002). On the efficiency of designs for linear models in non-linear regions and the use of standard designs for generalized linear models. Dissertation, Department of Statistics, Virginia Polytechnic Institute and State University.

International Journal of Statistics and Probability; Vol. 4, No. 4; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

A SAS Macro for Adaptive Spatial Sampling

Alan Ricardo da Silva[1] & Iracema Veiga Madeira Mauriz[1]

[1] Departamento de Estatistica, Universidade de Brasilia, Brazil

Correspondence: Alan Ricardo da Silva, Departamento de Estatistica, Universidade de Brasilia, Brazil. E-mail: alansilva@unb.br

Received: August 1, 2015 Accepted: August 25, 2015 Online Published: September 23, 2015

doi:10.5539/ijsp.v4n4p20 URL: http://dx.doi.org/10.5539/ijsp.v4n4p20

Abstract

This paper presents a SAS macro to estimate adaptive spatial sampling, which has been used to survey rare species. This technique is computationally difficult because of use of algorithms with GIS features such the creation of a grid, points inside polygons and contiguity. The results indicates that the SAS macro that was developed was capable of incorporating these GIS features, as well as estimating the parameters of the adaptive spatial sampling.

Keywords: adaptive sampling, spatial sampling, grid, SAS

1. Introduction

The purpose of sampling is to obtain information based on the results of a sample. According to Cochran (1977), sampling theory was developed to achieve the most efficient sampling, that is, to produce more accurate estimates with the lowest possible cost. Thus, the basic problem of any sampling procedure is to obtain reliable estimates of some characteristic of the population of interest, based on only part of this population.

A procedure that has been studied and tested in surveys of populations of rare species that display aggregated pattern distribution is adaptive sampling. In this kind of sampling, the selection of sampling units depends on observations made during the survey because if a criterion is met, the close sample is added to the initial sample. Thus, this type of sampling has advantages such as more extensive use of the sample and greater sampling intensity depending on the observations made during the survey; in addition, it can help find the local maximum (Thompson and Seber, 1996).

The adaptive sampling literature includes the following: algorithms that address the effects of mutations on the properties of folding RNA, the purpose of which is to decipher the principles of conduction and molecular evolution for the design of new molecules, in other words, these algorithms are for unbiased adaptive sampling that allows RNAmutants to sample regions of the mutational landscape that have not been fully addressed by previous techniques (Waldispühl and Ponty, 2011); designs for clinical study that focus on adaptation projects for two stage sample size re-estimation (Chang, 2008, 2009); mining applications that have a large amount of data, where random sampling may not be applicable due to the diffculty of determining an appropriate sample size (Domingo et al., 2002); and, finally, cases where the available algorithms for mining information on a large database are prohibitive due to computational constraints (time and memory) (Satyanarayana and Davidson, 2005).

In the case presented by Thompson (1990), another kind of adaptive sampling was considered, where the spatial distribution of aggregate data influences the formation of the sample selection of the data into clusters. This sample design, in which the procedure for the selection of units can be added to the initial sample based on an area and its spatial distribution, will be referred to from now on as adaptive spatial sampling.

Thus, adaptive spatial sampling provides a viable solution to the longstanding problem of estimating the abundance of rare populations and it has gained rapid acceptance in the natural and social sciences (Seber, 1986; Ramsey and Seber, 1992; Brown, 1994, 1996; Khan and Muttlak, 2002; Stein and Ettema, 2003; Sengupta and Sengupta, 2011; Jain and Chang, 2004; Thompson, 2011; Yu et al., 2012). However, adaptive procedures are more complicated to design and analyze, and computational implementations are few as a results of the complexity of the algorithms for spatial analysis (Thompson, 2011).

This implementation requires at least three steps: the development of a computational design for a regular grid; the selection of specific areas of the grid to identify which part of the grid the data are in; and identifying the neighbors of the selected areas: upper, lower, right and left. Thus, the objective of this work is to implement computationally adaptive spatial sampling in SAS software.

2. A Basic Outline of Adaptive Spatial Sampling

It has been observed that adaptive spatial sampling is performed in cluster sampling, because the data to be analyzed are divided into distinct subpopulations and have the geographical coordinates of a given area. Thus, the process of adaptive spatial sampling involves selecting certain areas, recording their geographical coordinates and defining the areas in which the data are located. Usually, these data are grouped in a particular area as shown in Figure 1(a), which characterize the clusters.

The first step is to draw a grid based on the geographical coordinates of the area to be analyzed, as shown in Figure 1(b). Then, some areas of the grid are selected by Simple Random Sampling (SRS) and it is determined whether there are data points within these areas, as shown in the blue areas of Figure 1(c). Next, the neighbors of these selected units are identified successively - top, bottom, right and left - until the selection criterion of the adaptive spatial sampling is exhausted. This adaptive spatial sampling is represented by one population sample (n) of the regular grid, as in Figure 1(d).

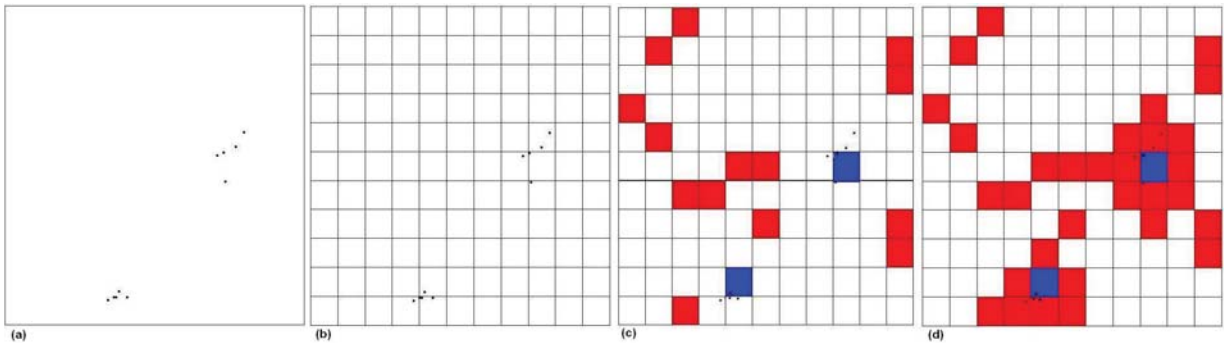


Figure 1. Steps of adaptive spatial cluster sampling

In summary, adaptive cluster sampling or simply adaptive spatial sampling refers to designs in which an initial set of units is selected by some probability sampling procedure and, whenever the variable of interest of a selected unit satisfies a given criterion, additional units in the neighborhood of that unit are added to the sample (Thompson, 1990). In the models considered in this paper, the initial sample can be selected by Simple Random Sampling with replacement SRS_R or without replacement SRS_{WR}.

2.1 Estimators

Classical estimators for the population mean are biased under an adaptive sampling design, in contrast with SRS. In this section two unbiased estimators for the population mean under an adaptive spatial sampling design will be addressed.

2.1.1 Estimators Using Initial Intersection Probabilities

This section shows an estimator based on a modification of a Horvitz-Thompson estimator (Thompson, 1990) and it is compared to the sample mean of the initial sample, given by

$$\bar{y} = \frac{1}{n_1} \sum_{i=1}^{n_1} y_i \tag{1}$$

where y_i is the variable in study of the unit i and n_1 is the initial sample size.

When an initial sample n_1 of units is selected by a SRS_{WR}, these units in the first sample are distinguished not as a result of replacement. However, the data itself may contain repeated observations if more than one unit in the cluster is selected in the initial sample. The unit i will be included in the final sample if any unit of A_i (including i itself), where A_i is a neighborhood of the point i, is selected as part of the initial sample, or if any unit of a network of which unit i is an edge unit is selected, where an edge unit is a neighborhood of the point i but without sample points.

Let m_i denote the number of units in A_i; N is the population size; and a_i, the total number of units in the network (neighbors of the selected grid: upper, lower, right and left), of which unit i is an edge unit. Note that if the unit

satisfies the criterion C, i.e., some data point is found inside the selected grid, then $a_i = 0$, but if the unit i does not satisfy this condition, then $m_i = 1$. The selection probability of unit i in either n_1 observations is $p_i = \frac{m_i + a_i}{N}$. The probability that unit i is included in the sample is given by (Thompson, 1990):

$$\Pi_i = P(I_i = 1) = 1 - \left[\binom{N - m_i - a_i}{n_1} \Big/ \binom{N}{n_1} \right] \tag{2}$$

When the selection of the initial sample is taken by SRS_R, repeated observations in the data can occur either because of possible repeated selections in the initial sample or the initial selection of more than one unit in the cluster. In this sample design, selection probability of unit i in either n_1 observation is $p_i = \frac{m_i + a_i}{N}$, and the probability of inclusion is given by (Thompson, 1990):

$$\alpha_i = 1 - (1 - p_i)^{n_1} = 1 - \left(1 - \frac{m_i + a_i}{N}\right)^n \tag{3}$$

If the values of Π_i are known for all sample units, one can use the Horvitz-Thompson estimator given by $\widehat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\Pi_i}$. However, although the values of m_i in Equation (2) for all units in the sample are known, only a few values of a_i are known. This means that i is a unit of edge somewhere in a cluster belonging to the sample, and thus, all clusters that this unit is related to do not need to be sampled. Thus, the value of a_i is unknown. To solve this problem, (Thompson, 1990) adopted the practice of dropping the value of a_i in Equation (2) and considering only the partial inclusion probability. Thus,

$$\Pi'_i = 1 - \left[\binom{N - m_i}{n_1} \Big/ \binom{N}{n_1} \right] \tag{4}$$

This probability Π'_i is now considered for n_1 networks instead of n_1 clusters and can be understood as the probability of the sample initial intercept A_i, the network for the unit i, to be used in the estimator. Thus, one obtains an unbiased estimator of the population mean based on the initial intersection probabilities as the following:

$$\widehat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i I'_i}{\Pi'_i} \tag{5}$$

where I'_i takes the value 1 (with probability Π'_i) if the initial sample intersects A_i, and 0 otherwise. In addition

$$\widehat{\mu}_{HT} = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\Pi_i} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i I_i}{\Pi_i} \tag{6}$$

where $y_1 \ldots y_n$ represent the n distinct values of units in the final sample and I_i has the value 1 when the unit is included in the sample and 0 otherwise.

Using the properties of mathematical expectation it turns out that the estimator of Equation (5) is unbiased,

$$E[\widehat{\mu}] = E\left[\frac{1}{N} \sum_{i=1}^{N} \frac{Y_i I'_i}{\Pi'_i} \right] = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i E(I'_i)}{\Pi'_i} = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i \Pi'_i}{\Pi'_i} = \frac{1}{N} \sum_{i=1}^{N} Y_i = \mu \tag{7}$$

The classical estimator of the population mean under adaptive spatial sampling design is a biased estimator as follows:

$$E[\overline{y}] = E\left[\frac{1}{n_1} \sum_{i=1}^{n_1} y_i \right] = \frac{E\left(\sum_{i=1}^{N} y_i I'_i \right)}{n_1} = \frac{\sum_{i=1}^{N} Y_i E(I'_i)}{n_1} = \frac{\Pi'_i N \mu}{n_1} \neq \mu \tag{8}$$

To facilitate the analysis of Equation (5) it is more convenient to rewrite it in terms of distinct networks because the probability of intersection Π'_i is the same (also called α_k) for each unit i in the kth network. Thus,

$$\alpha_k = 1 - \left[\binom{N - x_k}{n_1} \Big/ \binom{N}{n_1} \right] \tag{9}$$

Similar to the equations of the probability of inclusion and as p_{jk} is the probability that kth and jth networks do not intersect, so

$$p_{jk} = P(J_j \neq 1 \cap J_k \neq 1) = \binom{N - x_j - x_k}{n_1} \Big/ \binom{N}{n_1} \tag{10}$$

where x_j is the number of units in the k-th network and J_k is the initial sample intersect of the kth network and takes the value 1 (with probability α_k) and 0 otherwise.

Using Equations (9) and (10) we obtain α_{jk} (the probability of intersection of the kth and jth networks) as

$$\begin{aligned}
\alpha_{jk} &= \alpha_j + \alpha_k - (1 - p_{jk}) \\
&= 1 - \left[\binom{N - x_j}{n_1}\right] \Big/ \binom{N}{n_1} + 1 - \left[\binom{N - x_k}{n_1} \Big/ \binom{N}{n_1}\right] - \left[1 - \binom{N - x_j - x_k}{n_1} \Big/ \binom{N}{n_1}\right] \\
&= 1 - \frac{\binom{N-x_j}{n_1}}{\binom{N}{n_1}} + 1 - \frac{\binom{N-x_k}{n_1}}{\binom{N}{n_1}} - 1 + \frac{\binom{N-x_{j-x_k}}{n_1}}{\binom{N}{n_1}} \\
&= 1 - \left[\binom{N - x_j}{n_1} + \binom{N - x_k}{n_1} - \binom{N - x_j - x_k}{n_1}\right] \Big/ \binom{N}{n_1}
\end{aligned} \tag{11}$$

Therewith,

$$\widehat{\mu} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k^* J_k'}{\alpha_k} = \frac{1}{N} \sum_{k=1}^{K} \frac{y_k^*}{\alpha_k} \tag{12}$$

where y_k^* is the sum of the y-values for kth network, K is the total number of distinct networks in the population, and k is the number of distinct networks in the sample.

Let $z_k = y_k^*/\alpha_k$, $y_k^* = \sum_{i=1}^{N} y_i$ $\Pi_K = \alpha_k$ and Π_{jk}. From the properties of mathematical expectation, variance, covariance and the definitions above, one can obtain the expected value and the variance of Equation (5) by the following:

$$E[\widehat{\mu}] = \frac{1}{N} \sum_{k=1}^{K} z_k E(J_k) = \frac{1}{N} \sum_{k=1}^{K} z_k \alpha_k = \frac{1}{N} \sum_{k=1}^{K} y_k^* = \frac{1}{N} \sum_{i=1}^{N} y_i^* = \overline{y} = \frac{\tau}{N} = \mu \tag{13}$$

$$\begin{aligned}
var[\widehat{\mu}] &= var\left[\frac{1}{N} \sum_{k=1}^{K} z_k J_k\right] = \frac{1}{N^2} \left[\sum_{k=1}^{K} z_k J_k + \sum_{j=1}^{K} \sum_{j \neq k} cov\left(z_j J_j z_k J_k\right)\right] \\
&= \frac{1}{N^2} \left[\sum_{j=1}^{K} z_j^2 \Pi_j (1 - \Pi_k) + \sum_{j=1}^{K} \sum_{j \neq k} z_j z_k \Pi_{jk} - \Pi_i \Pi_k\right] \\
&= \frac{1}{N^2} \left[\sum_{j=1}^{K} \sum_{k=1}^{K} z_j z_k (\Pi_{jk} - \Pi_j \Pi_k)\right] \\
&= \frac{1}{N^2} \left[\sum_{j=1}^{K} \sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_j \alpha_k}\right)\right]
\end{aligned} \tag{14}$$

and an unbiased estimator of the variance of Equation (14) is:

$$\begin{aligned}
\widehat{var}[\widehat{\mu}] &= \sum_{j=1}^{K}\sum_{k=1}^{K} z_j z_k J_j J_k \left(\frac{\Pi_{jk} - \Pi_j \Pi_k}{\Pi_{ij}}\right) \\
&= \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_{jk}\alpha_j\alpha_k}\right) J_j J_k\right] \\
&= \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} y_j^* y_k^* \left(\frac{\alpha_{jk}}{\alpha_{jk}\alpha_j\alpha_k} - \frac{1}{\alpha_{jk}}\right)\right] \\
&= \frac{1}{N^2}\left[\sum_{j=1}^{K}\sum_{k=1}^{K} \frac{y_j^* y_k^*}{\alpha_{jk}} \left(\frac{\alpha_{jk}}{\alpha_j\alpha_k} - 1\right)\right] \qquad (15)
\end{aligned}$$

Another known estimator for adaptive spatial cluster sampling is one that uses the expected number of initial intersection as follows.

2.1.2 Estimators Using the Expected Number of Initial Intersection

The estimator given by Equation (5) can be rewritten as:

$$\widetilde{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i \frac{f_i}{E[f_i]} \qquad (16)$$

where f_i represents the number of units in the initial sample that fall in network A_i, which includes the unit i; N is the number of regular grids. If during the estimation process the units edge of clusters is ignored, f_i would be interpreted as the number of times the ith unit of the final sample appears in the estimator. Then one realizes that $f_i = 0$ if no units in the initial sample intersect A_i.

The estimator in Equation (16) is unbiased because

$$E[\widetilde{\mu}] = E\left[\frac{1}{N}\sum_{i=1}^{N} y_i \frac{f_i}{E[f_i]}\right] = \frac{1}{N}\sum_{i=1}^{N} E(y_i)\frac{E[f_i]}{E[f_i]} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \mu \qquad (17)$$

Because m_i is the number of units on the network to which i belongs, using the Horvitz-Thompson estimator, another unbiased estimator can be found: As f_i units are selected from m_i units in A_i, f_i follows a hypergeometric distribution with the following parameters: (N, m_i, n_1) (Thompson, 1991). Thus, $E[f_i] = \frac{n_1 m_i}{N}$ and substituting the expected value in Equation (16) we obtain the following:

$$\widetilde{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i \frac{f_i}{\frac{n_1 m_i}{N}} = \frac{N}{N}\sum_{i=1}^{N} \frac{y_i f_i}{n_1 m_i} = \frac{1}{n_1}\sum_{i=1}^{N} \frac{y_i f_i}{m_i} \qquad (18)$$

To find the variance of the estimator of Equation (18), the approach in terms of n_1 networks being connected is used, although this is not necessarily distinct. Because m_i has the same value for all units in A_i and w_i is the average of m_i observations A_i (Thompson, 1990), then

$$\widetilde{\mu} = \frac{1}{n_1}\sum_{i=1}^{n_1} \frac{1}{m_i}\sum_{j\epsilon A_i} y_j = \frac{1}{n_1}\sum_{i=1}^{n_1} w_i = \overline{w} \qquad (19)$$

Thus, $\widetilde{\mu}$ is the sample mean obtained by taking a selection of an SRS of size n_1 of a population of w_i values rather then y_i values. Because $w_i = \overline{v}_k$ is the same for each unit in the kth network, where $\overline{v}_k$ is the mean of the y-values in β_k, there are x_k units in the kth network and B_K is a set of units in the kth network; thus,

$$E(\widetilde{\mu}) = E(\overline{w}) = E\left[\frac{1}{N}\sum_{i=1}^{N} w_i\right] = \frac{1}{N}\sum_{k=1}^{K} x_k \overline{v}_k = \frac{1}{N}\sum_{k=1}^{K}\sum_{i\epsilon B_k} y_i = \mu \qquad (20)$$

From the equations of cluster sampling in two stages, we obtain an unbiased estimator of this variance:

$$var[\widetilde{\mu}] = var\left[\frac{1}{N}\sum_{i=1}^{N} w_i\right] = \frac{N-n_1}{Nn_1(N-1)}\sum_{i=1}^{N}(w_i-\mu)^2 = \frac{\sigma^2}{n_1}\left(1-\frac{n}{N}\right) \tag{21}$$

where $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(w_i-\mu)^2$.

$$\widehat{var}[\widetilde{\mu}] = \frac{N-n_1}{Nn_1(n_1-1)}\sum_{i=1}^{n_1}(w_i-\widetilde{\mu})^2 \tag{22}$$

As in some populations a priori information may be known; i.e., where aggregations occur, one can use the technique to reduce the stratified sample variance estimators. To do this, we can use stratified adaptive spatial sampling, where the population is divided into strata and the number of units are sampled for each stratum that is used.

3. Stratified Adaptive Spatial Cluster Sampling

In the case of adaptive spatial sampling techniques, one must also know the geographic coordinates of the selected area. Thereafter, one stratifies the area using a priori information and draws the grid throughout the selected area (including stratified areas) through their respective locations as indicated in Figure 2 (a). After that, a sample is selected using *S RS*, as in Figure 2 (b). Then, the units of interest are identified, as in Figure 2 (c). Finally, one successively adds the neighbors of selected areas - upper, lower, right and left - until the selection criterion of the adaptive sample is exhausted, as in Figure 2 (d).

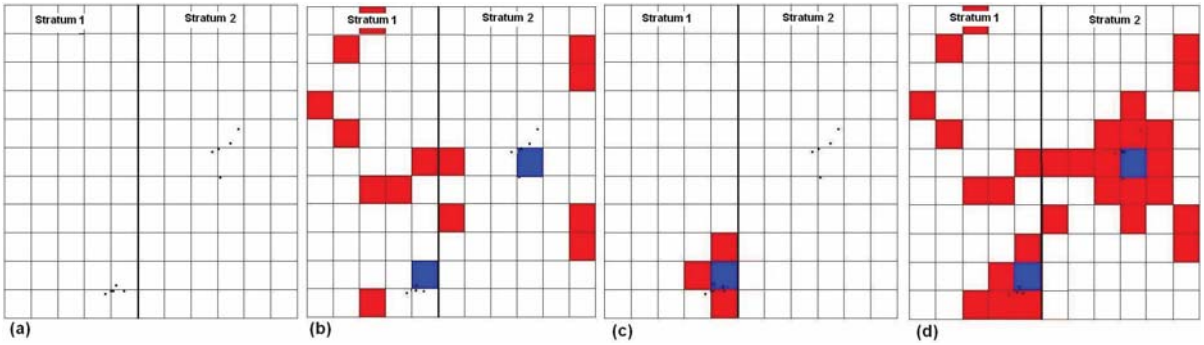


Figure 2. Steps of stratified adaptive spatial cluster sampling

In the estimators to be considered, the initial sample can also be selected by SRS_R or SRS_{WR}. The next section will show three unbiased estimators for this kind of sampling.

3.1 Estimators

Suppose that the population total of N units is partitioned into L stratum, with n_h units in the hth stratum ($h = 1, 2, \ldots, L$). Define unit (h, i) as the ith unit in the hth stratum with associated y-value y_{hi}. This process begins with a SRS of n_h units that is taken from stratum L, and we now define $n_0 = \sum_{h=1}^{L} n_h$ to be the initial total sample size. From this, the clusters begin to have neighbors added according to the condition set C (Thompson and Seber, 1996).

3.1.1 Estimators Using Initial Intersection Probabilities

Using the full adaptive sample, the first estimator we can consider is given by (5) based on the initial intersection probabilities, we obtain

$$\widehat{\mu}_{st} = \frac{1}{N}\sum_{k=1}^{K}\frac{y_k^* J_k}{\alpha_k} \tag{23}$$

where the K distinct networks are labeled $(1, 2, \ldots, k)$ without regard for stratum boundaries, J_k equals 1 (with probability α_k) if the initial sample size n_0 intersects network k, and 0 otherwise and, finally, y_k^* is the sum of the y-values for the network k.

To derive α_k it is necessary to consider the probabilities of intersecting network k with the initial samples in each strata. Therefore, we define x_{hk} as the number of units in stratum h that lie in network k. This number assumes the

value 0 if the network k lies totally outside stratum h. If the network straddles a boundary, we ignore the network units that lie outside *stratum h* in the definition of x_{hk}. Thus, with this definition of x_{hk}, we obtain

$$\alpha_k = 1 - \left[\prod_{h=1}^{L} \frac{\binom{N_h - x_{hk}}{n_h}}{\binom{N_h}{n_h}} \right] \tag{24}$$

The variance of the estimator of the unbiased average, defined as the probability of the initial sample intercede network in k and k', is obtained in the following way:

$$\alpha_{kk'} = 1 - (1 - \alpha_k) - (1 - \alpha_{k'}) + \left[\prod_{h=1}^{L} \frac{\binom{N_h - x_{hk} - x_{hk'}}{n_h}}{\binom{N_h}{n_h}} \right] \tag{25}$$

Because $\alpha_{kk} = \alpha_k$ and from the variance properties, it follows that

$$var[\widehat{\mu}_{st}] = \frac{1}{N^2} \sum_{k=1}^{K} \sum_{k'=1}^{K} y_k^* y_{k'}^* \left(\frac{\alpha_{kk'} - \alpha_k \alpha_{k'}}{\alpha_k \alpha_{k'}} \right) \tag{26}$$

and an unbiased estimator for this variance is given by

$$\widehat{var}[\widehat{\mu}_{st}] = \frac{1}{N^2} \sum_{k=1}^{K} \sum_{k'=1}^{K} y_k^* y_{k'}^* \left(\frac{\alpha_{kk'} - \alpha_k \alpha_{k'}}{\alpha_{kk'} \alpha_k \alpha_{k'}} \right) I_k I_{k'} \tag{27}$$

Another estimator for this kind of sampling is to use the expected number of the initial intersection, which will be explained in the next section.

3.1.2 Estimators Using the Expected Number of the Initial Intersection

Let A_{hi}, the network that contains the unit (h, i), u_{hi}, and A_{ghi} be part of A_{hi} stratum g. Suppose that f_{ghi} is the number of units from the initial sample in stratum g that fall in A_{ghi}, and let m_{ghi} be the number of units in A_{ghi}. Then, the number of units of an initial sample n_0 units is A_{hi} given by (Thompson and Seber, 1996),

$$f_{\cdot hi} = \sum_{g=1}^{L} f_{ghi} \tag{28}$$

From Equation (16) one obtains the estimator for the mean

$$\widetilde{\mu}_{st} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} y_{hi} \frac{f_{\cdot hi}}{E[f_{\cdot hi}]} \tag{29}$$

As in Equation (7) and from the properties of expectation and variance, this estimator is unbiased.

As with f_i in Equation (16), f_{ghi} follows a hypergeometric distribution with parameters (N_g, m_{ghi}, n_g). Therefore, it is known that $E[f_{ghi}] = \frac{n_g m_{ghi}}{N_g}$ and $E[f_{\cdot hi}] = \sum_{i=1}^{L} \frac{n_g}{N_g} m_{ghi}$ (Thompson, 1991). Thus,

$$\widetilde{\mu}_{st} = \frac{1}{N} \sum_{h=1}^{L} y_{hi} \frac{f_{\cdot hi}}{\sum_{i=1}^{L} \frac{n_g}{N_g} m_{ghi}} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \left(y_{hi} \sum_{g=1}^{L} f_{ghi} / \sum_{g=1}^{L} \frac{n_g}{N_g} m_{ghi} \right) \tag{30}$$

where f_{ghi} represents the number of units in the initial sample that is at the intersection of stratum g with the network drive to which the unit u_{hi} belongs.

If there is a match to add the same neighbors, we obtain an estimator of the independent stratum combined with the estimator with weights, providing an estimator of the population mean as Equation $E[\overline{y}_{st}] = E\left(\sum_{h=1}^{L} \frac{N_h}{N} \overline{y}_h\right)$. This characteristic aggregation of equal neighbors generates a loss of efficiency, a more efficient system would allow groups to overlap the boundaries of the strata (Thompson, 1991).

So, to find the variance estimator of the mean, we use Equation (19) to rewrite $\widetilde{\mu}_{st}$ in terms of the weights of the sample means. For this, relate the observations to the intercept of the initial sample networks. Thus, the term $y_{hi}f_{.hi}$ means that A_{hi} is intersected f_{hi} times by the initial sample, so that $\widetilde{\mu}_{st}$ represents a weighted sum of all units in all the networks corresponding to the initial sample, with some networks being repeated. Because the weight $E[f_{.hi}]$ is the same for each unit in A_{hi}, we have

$$\widetilde{\mu}_{st} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{n_h}\frac{1}{E[f_{.hi}]}\sum_{(h',i')\epsilon A_{hi}} y_{h'i'} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{n_h}\frac{Y_{hi}}{E[f_{.hi}]} \tag{31}$$

where Y_{hi} is the sum of yth observations in A_{hi}.

Let $\overline{w}_h = \sum_{i=1}^{n_h}\frac{w_{hi}}{n_h}$ and $w_{hi} = \frac{n_h Y_{hi}}{N_h E[f.hi]}$; then, another way to rewrite Equation (31) is

$$\widetilde{\mu}_{st} = \sum_{h=1}^{L}\frac{N_h}{N}\overline{w}_h = \frac{1}{N}\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h} w_{hi} \tag{32}$$

where $w_{hi} = \frac{Y_{hi}}{\sum_g m_{ghi}}$; when $\frac{n_h}{N_h}$ have the same value for all strata. Thus, Equation (32) represents a stratified sample mean from a stratified random sampling without replacement, with the interest of w_{hi} variable.

So, the variance estimator for the mean is given by

$$var[\widetilde{\mu}_{st}] = \frac{1}{N^2}\sum_{h=1}^{L} N_h(N_h - n_h)\frac{\sigma_h^2}{n_h} \tag{33}$$

where σ_h^2 represents the stratum population variance, that is,

$$\sigma_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(w_{hi} - \overline{W}_h)^2 \tag{34}$$

where $\overline{W}_h = \frac{\sum_{i=1}^{n_h} w_{hi}}{n_h}$ is the stratum population mean.

An unbiased estimator of variance of the mean (33) can be obtained by replacing σ_h^2 by sample variance, $s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(w_{hi} - \overline{w}_h)^2$.

3.1.3 Estimators that Ignore Units Added through Crossing Boundaries

According to Thompson (1991), the estimator that ignores units added through crossing stratum boundaries is given by the following:

$$\mu_{st}^{''} = \sum_{h=1}^{L}\frac{N_h}{N}\widetilde{\mu}_h = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(y_{hi}\sum_{g=1}^{L}\frac{N_g}{n_g}f_{ghi}/\sum_{g=1}^{L} m_{ghi}\right) \tag{35}$$

where $\widetilde{\mu}_h = \sum_{i=1}^{n_h}\frac{w_{hi}^{''}}{n_h}$ and $w_{hi}^{''}$ is the total of the y-values in the intersection of the stratum h with A_{hi} divided by the number of units in the intersection; i.e., this value represents the network mean for that part of the network A_{hi} in stratum h.

The mathematical expectation of $\mu_{st}^{''}$ is given by

$$E[\mu_{st}^{''}] = \sum_{h=1}^{L}\frac{N_h}{N}\mu_h = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi} = \mu \tag{36}$$

The variance $var[\mu_{st}^{''}]$ is given by

$$var[\widetilde{\mu}_{st}^{''}] = \frac{1}{N^2}\sum_{h=1}^{L} N_h(N_h - n_h)\frac{\sigma^2}{n_h} \tag{37}$$

where the stratum population variance is $\sigma_h^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(w_{hi}^{''} - \overline{W}_h)^2$ and the stratum population mean is $\overline{W}_h = \sum_{i=1}^{N_h}\frac{w_{hi}}{N_h}$. The sample estimate is given by (32) replacing σ_h^2 by $s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(w_{hi} - \overline{w}_h)^2$.

4. SAS Macros

The SAS Macros basically use the IML Procedure and GMAP and SQL Procedures. The computational implementation of adaptive spatial sampling requires four steps: 1) development of the computational design of regular grids (Figure 3(a)); 2) selection of specific areas of the grid, that is, identifying a sample and determining in which part of the grid the data are located (Figure 3(b)); 3) identification of the neighbors of the selected areas: upper, lower, right and left (Figure 3(c)(d)); 4) calculation of the parameters.

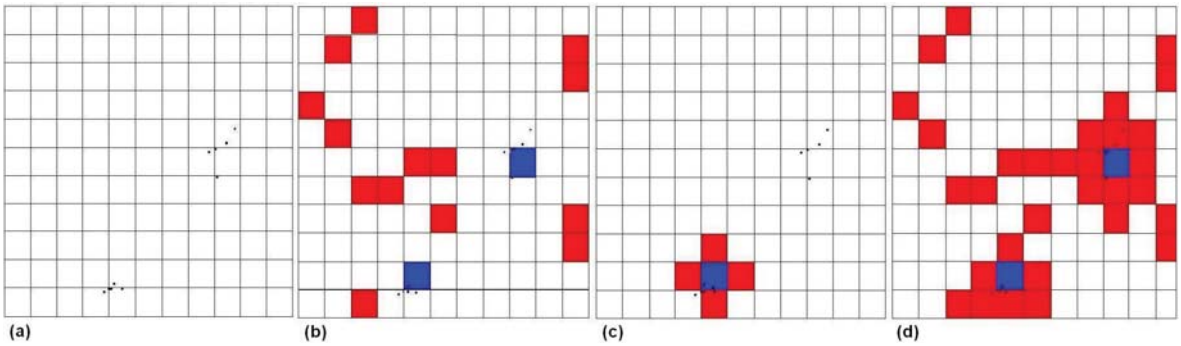


Figure 3. Steps of adaptive spatial sampling

4.1 Drawing a Regular Grid

To create regular grids it is necessary to create four points with coordinates entered clockwise (lines of Table 1) or counterclockwise.

Table 1. Coordinates of a square

Reference	Values	*Points*
1	**(Min, Min)**	(0,0)
2	**(Min, Max)**	(0,1)
3	**(Max, Max)**	(1,1)
4	**(Max, Min)**	(1,0)

In the case of a square, beginning with the clockwise points, i.e., the reference points in the following order: 1, 2, 3, 4, the polygon appears to be like Figure 4 (Square). If that order is not followed, the result is a distorted polygon, as shown in Figure 4 (Distorted Polygon).

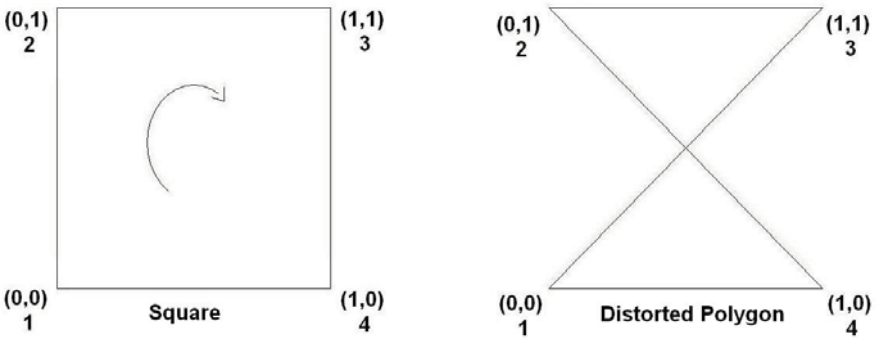


Figure 4. Square and distorted polygon.

As one wishes to draw a grid on a field of study, it is necessary to know the upper and lower boundaries of the region, i.e., the minimum and maximum coordinates of the *y* axis (latitude) and minimum and maximum coordinates of the *x* axis (longitude). The definition of the size of each polygon is given by **%grid** macro:

```
%grid(minx =, maxx =, miny =, maxy =, dim =, anno =, printN = YES );
```

where the parameters are: **MINX =** the minimum value of the **x** coordinate; **MAXX=** the maximum value of the **x** coordinate. Similarly, for the **y** coordinates we have: **MINY=** and **MAXY=**. Another parameter of this macro is the size of the square drawn given by **DIM=** (for instance, if **DIM=** 20, $20^2 = 400$ squares will be created). Finally, the last two parameters, **ANNO=** and **PRINTN = YES**, indicate the dataset with the location of samples and whether the numbering of each square will be printed using the command **YES**(Figure 5 (a)) or **NO** (Figure 5 (b)), respectively.

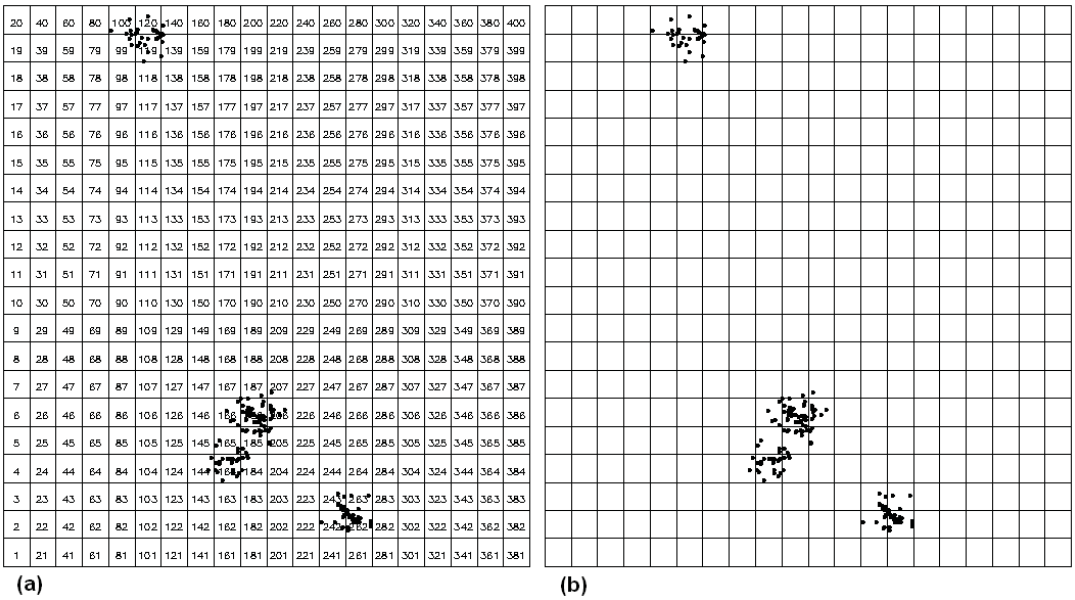


Figure 5. Regular grid for $N = 400$.

Next, we define the element, **id** to the coordinates of the square. Thus, the first square has the points $(0, 0), (0, 1), (1, 0), (1, 1)$, **id = 1** and so on. This is achieved by joining the table with the coordinates with the table set out below:

```
data id&dim;
 do id=1 to &dim*&dim;
  do i=1 to 4;
   output;
  end;
 end;
run;
```

This numbering starts from a unit numeric value and goes to the value of N to count the vertical direction, starting from left to right. In the case of Figure 5, it is found that the size of the square is $N = 20 \times 20 = 400$, varying the **id** from 1 to 400.

The next step is to make the selection of specific areas of the grid; i.e., the grids are drawn by SRS, and if there are samples inside a grid, it is selected by its neighbors.

4.2 Selection of Specific Areas of the Grid

The samples to be drawn in adaptive spatial sampling correspond to the polygons of the grid. This selection can be done by a generator corresponding to the number of grid squares of random numbers. In SAS, this can be done by PROC SURVEYSELECT, where a SRS is obtained with a seed value of size n given by the variable **SEED**. The parameter **OUT** indicates where the sample will be stored.

```
proc surveyselect data=&data sampsize=&n out=&saida seed=&seed noprint;
run;
```

The identification of the neighbors of the selected areas in the next section involves three concepts: check point inside the polygon; definition of the neighbors; and identification of neighboring polygons.

4.3 Identification of the Neighbors

The selection of neighbors is the trickiest and the most important part of the adaptive spatial sampling technique, as it is from the selected grids that the process of adapting the sample areas begins.

4.3.1 Checking if the Point is Inside the Polygon

The determination of points inside the polygon is shown in Figure 6. The main idea consists in making a radius from the selected point p to infinity in any direction and computing the amount of times that this line passed through the edges of the polygon. Thus, if the number of crossings is odd, the point is inside the polygon (Kunigami, 2010).

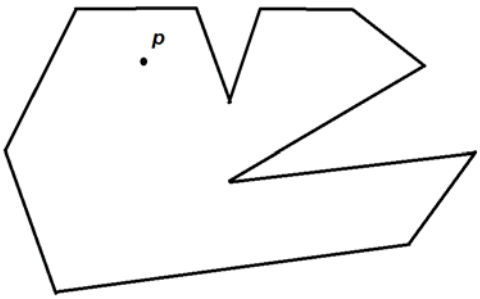


Figure 6. Point inside polygon.

The selection of these points within the regular grid is done by **%ginside** macro.

```
%ginside(map = , id = , where = , data = , out = );
```

where the elements of the macro are: **MAP =** dataset containing the coordinates of the study area; **ID=** name of the ID variable that defines the polygon; **WHERE =** selecting a specific point to check if it is inside the polygon; **DATA =** dataset containing the points to be checked if they are inside the polygon; **OUT =** dataset in which the points inside the polygons will be stored.

4.3.2 Definition of the Neighbors

The definition of the neighbors is as follows: The neighborhood is a set of squares that are added to the same sample if the grid satisfies the same condition of containing elements of interest in the selected square. We define the neighborhood of type ROOK as the polygons that share more than one point in common, in this case the square: up, down, right and left, as shown in Figure 3(c), and the neighborhood of type QUEEN as the polygons that share at least one point in common, i.e., the neighborhood ROOK adding the corners.

The **%neighborhood** macro is given by:

```
%neighborhood(id = , pt = , map = , anno = , out = , type = ROOK );
```

where the parameters are: **ID=** name of the ID variable that defines the polygon; **PT=** ID of the grid in which the neighbors will be defined; **MAP=** the dataset containing the coordinates of the study area; **ANNO=** the dataset containing the coordinates of the samples; **OUT=** the dataset in which the neighbors will be stored; **TYPE=** indicates that the pattern of selection of the neighbors is of type **ROOK** (default) or **QUEEN**.

4.3.3 Identifying Neighboring Polygons

The identification of neighboring polygons is generated by combining the identification of neighbors with the points inside the polygon; this is the final sample of adaptive spatial sampling.

Finally, given the base with the selected units and their respective score points, the next step is to compute the estimators presented in Section 2.

4.4 Estimators of Adaptive Spatial Sampling

In this section the formulas for the estimators of Section 2 were implemented. Thus, for adaptive spatial sampling we have implemented the estimator of the mean (**u1**), Equation (19), the estimator of the variance of the estimator

of the mean (**varu1**), Equation (21) the estimator of the total (**Totu1**) given by multiplying the number of grids by the estimator **u1**, i.e., **Totu1 = NN× u1**, and the estimator of the variance of **Totu1**, named **TotVaru1**.

The **%as** macro computes the estimators and automatically uses the macros presented previously.

```
%as(data = , n = , sample = , out = , strata = , seed = , map = , id = ,
    anno = , typen = ROOK , printN = YES );
```

where the parameters are the same as those presented earlier, and **n=** the sample size and **SAMPLE=** a dataset containing a predefined sample.

5. Illustration

5.1 An Example of Adaptive Spatial Cluster Sampling

Thompson (1990) presents an example of how adaptive spatial sampling works and compares the results obtained from the SRS estimator and from SRS with adaptive spatial sampling, which is given by changing the denominator of Equation (19) to the denominator of Equation (8). This example could represent a reserve of animals that are grouped (as herds of elephants) or deposits of minerals (such as gold, diamond, iron) spread over large areas.

Initially, a regular grid is drawn on the area to be sampled, and then, n units (squares) are selected by the SRS method. In this example, the initial sample consists of 10 units (total squares on the grid in red) selected in a total of $N = 400$ units (representing the total number of regular square grids, where each side has a length of 20, or $20 \times 20 = 400$), with a total of 190 points (Figure 7 (c)).

Selecting the neighbors (by the ROOK methodology) of the initial units containing at least one unit in the initial sample, we obtain the final sample, as shown in Figure 7(d). The upper unit has an element that intersects with the network $m_1 = 6$ units, containing a total of $y_1^* = 36$ units of interest. Another point in which there is unity within the polygon that intersects the network $m_2 = 11$ units and contains $y_2^* = 107$ units. For the other 8 units of the initial sample, the values are $y_i = 0$ and $m_i = 1$.

There are also 20 edge units that are not used in calculating the estimates; these are selected for adaptive selection, but they do not contain units of interest. In Figure 7(d) networks within the two groups that are described as being adaptively added are in the color red.

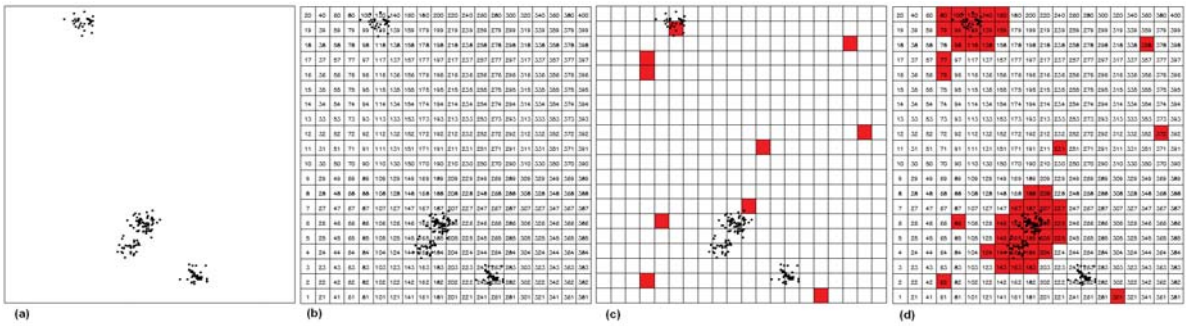


Figure 7. Example of adaptive cluster sampling.

For $w_1 = \frac{36}{6} = 6$ objects per unit, for $w_2 = \frac{107}{11} = 9.727$ and for the remaining $w_i = 0$, one can calculates the values for the mean estimators, $\widetilde{\mu}$, and from the total of the adaptive sampling:

$$\begin{aligned}
\widetilde{\mu} &= \frac{1}{10}\left[\frac{36}{6} + \frac{107}{11} + \left(\frac{0}{1}\right) + \ldots + \left(\frac{0}{1}\right)\right] = 1.573 \\
N\widetilde{\mu} &= 400 \times 1.573 = 629 \\
\widehat{var}[\widetilde{\mu}] &= \frac{(400-10)}{400(10)(10-1)}[(6-1.573)^2 + \ldots + (0-1.573)^2] = 1.147 \\
N^2\widehat{var}[\widetilde{\mu}] &= 400^2 \times 1.147 = 183,520
\end{aligned}$$

For SRS, where N is the total number of square areas selected and n is the number of selected squares of the initial

sample, one obtains the values for the estimators of the mean $\overline{y}$ and for the total $N\overline{y}$:

$$\begin{aligned}\overline{y} &= \frac{11+1}{10} = 1.2\\ N\overline{y} &= 400 \cdot 1.2 = 480\\ \widehat{var}[\overline{y}] &= 1.165\\ N^2\widehat{var}[\overline{y}] &= 186,506\end{aligned}$$

For the 45 units, which includes the 25 edge units of the final sample, the values for the estimators of the average $\overline{y}_{AD}$ for adaptive SRS are calculated; that is, the amount of SRS Equation in Equation (19) is used. Thus,

$$\begin{aligned}\overline{y}_{AD} &= \frac{143}{45} = 3.178\\ N\overline{y}_{AD} &= 400 \times 3.178 = 1,271\\ \widehat{var}[\overline{y}_{AD}] &= 1.004\\ N^2\widehat{var}[\overline{y}_{AD}] &= 400^2 \cdot 1.004 = 160,687\end{aligned}$$

Table 2 presents the estimates found and it appears that the variance, mean and total sampling of the adaptive SRS is the lowest compared to the others. However, their estimated average is higher, because there is a bias when using the estimator of SRS in this sample. Comparing the ratio of the variances $\overline{y}_{AD}$ and $\widetilde{\mu}$, we observe that there is a reduction of 13% in this value. Thus, given that it has a total of 190 points and the actual population mean is $\mu = \frac{190}{400} = 0.475$, adaptive spatial sampling in this case was very close to the SRS; however as will be seen later, adaptive spatial sampling varies much less when N varies.

The computational output of the program implemented in SAS software for this example is given in Figure 8 (the results are the same as in Thompson (1990)). Thus, there is the number of observations ($n = 10$), population size ($N = 400$), the estimators for the mean and total and their respective variances in the three cases analyzed: adaptive spatial cluster sampling, SRS and adaptive SRS sampling (biased).

Adaptive Cluster Sampling

Number of Observations: 10

Population Size: 400

Adaptive Cluster Sampling

Statistics

Mean	Var of Mean	Sum	Var of Sum
1.5727273	1.1470888	629.09091	183534.21

Simple Random Sampling

Statistics

Mean	Var of Mean	Sum	Var of Sum
1.2	1.1656667	480	186506.67

Adaptive Simple Random Sampling (biased)

Statistics

Mean	Var of Mean	Sum	Var of Sum
3.1777778	1.0042994	1271.1111	160687.9

Figure 8. Output of adaptive spatial cluster sampling.

Table 2. Table of comparison of the estimators of adaptive sampling, SRS and adaptive SRS

Estimator	$\widetilde{\mu}$	$\bar{y}$	$\bar{y}_{AD}$
Mean	1.57	1.20	3.17
Total	629	480	1,271
Variance of the mean	1.147	1.165	1.004
Variance of the total	183,520	186,506	160,687

5.2 Comparison between Different Population Sizes

In this section, interference variation for the total areas (grids) in the estimates will be checked, i.e., the variation in N. Thus we have simulated different sizes of squares and the initial samples were set to have the same sample as the sample (Thompson, 1990). Thus, we obtained the results in Table 3, with the respective values of the estimators of the mean and their estimated variances.

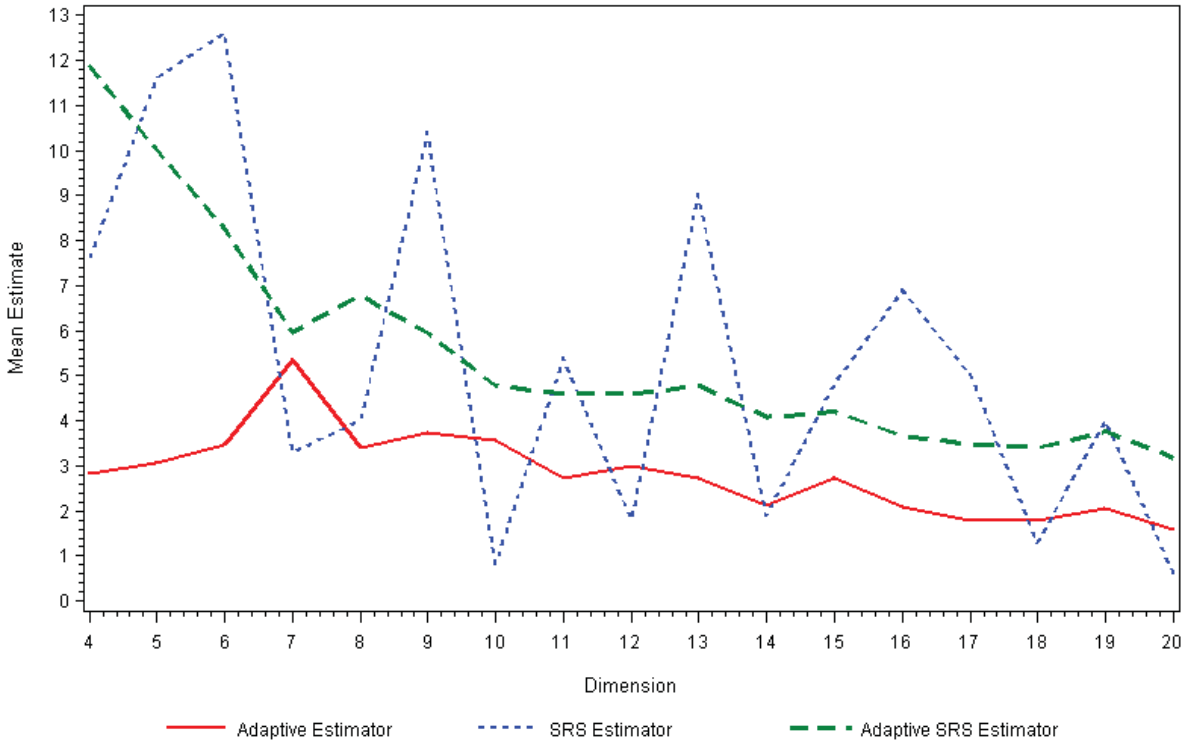


Figure 9. Analysis of the mean when the population size increases (ROOK).

Figure 9 shows how the estimated average is influenced in the case of population variation. We can see that adaptive spatial cluster sampling (solid green line) had the lowest average interference with the change in population, undergoing a decrease with an increase in the size of the regular grid, except when $N = 7$. SRS (blue dotted line) underwent a large change throughout the process. In addition, adaptive SRS sampling - $SRSAD$ (dashed red line) - shows a decrease with a slight increase when $N = 8$.

Table 3. Comparison between the estimators of Mean: adaptive sampling, SRS, adaptive SRS (ROOK)

Matrix	N	$\bar{\mu}$	$\widehat{var}(\bar{\mu})$	$\bar{y}$	$\widehat{var}(\bar{y})$	$\bar{y}_{AD}$	$\widehat{var}(\bar{y}_{AD})$
4x4	16	2.82	1.35	7.60	10.30	11.87	0
5x5	25	3.08	3.43	11.60	42.33	10.00	6.06
6x6	36	3.48	5.99	12.60	62.66	8.26	6.64
7x7	49	5.36	11.57	3.30	3.89	5.96	5.50
8x8	64	3.40	4.70	4.00	7.52	6.78	4.28
9x9	81	3.72	5.41	10.40	47.76	5.93	3.88
10x10	100	3.58	5.13	0.80	0.27	4.76	4.01
11x11	121	2.72	3.01	5.40	13.54	4.61	2.50
12x12	144	2.98	3.85	1.80	1.42	4.61	3.13
13x13	169	2.73	3.10	9.00	36.88	4.76	4.51
14x14	196	2.13	2.31	1.90	1.95	4.08	1.83
15x15	225	2.73	3.22	4.80	10.82	4.20	1.77
16x16	256	2.09	1.91	6.90	22.69	3.67	1.71
17x17	289	1.79	1.44	5.00	15.55	3.49	1.42
18x18	324	1.79	1.45	1.30	0.73	3.40	1.06
19x19	361	2.06	2.03	4.00	7.69	3.76	1.21
20x20	400	1.57	1.15	1.20	1.16	3.18	1.00

Figure 10 represents how the estimate of the variance of the average is influenced in the case of variation in the population. It is apparent that the variance in adaptive sampling is less affected by population size, having a range of 0 to 15, while SRS ranges from 10 to 75 and $SRSAD$ from 0 to 15.

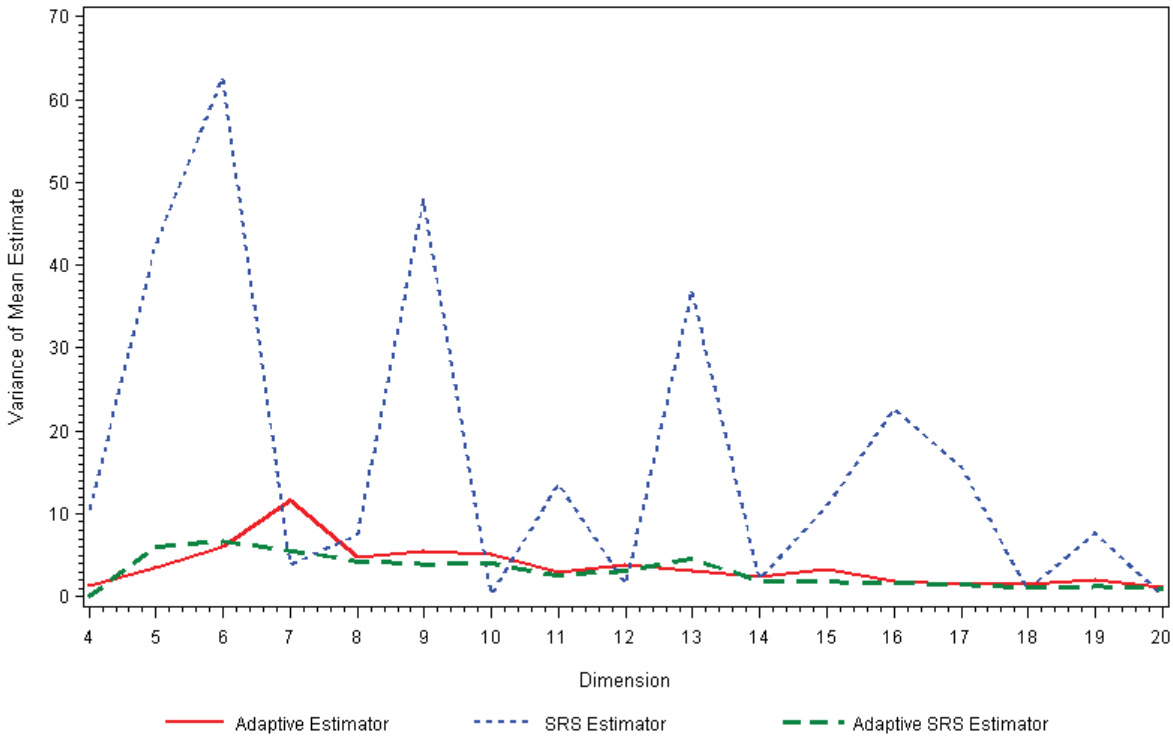


Figure 10. Analysis of the variance when the population size increases (ROOK).

Similarly, we obtain the results in Table 4 for the estimators of the Total. Figure 11 shows the behavior of the total estimator when the population size varies. For this case, the estimator of the total adaptive sampling - $N\mu$ - is

the one with less variation when compared with the other two ($N\bar{y}$ and $N\bar{y}_{AD}$). The variation of the estimator $N\bar{y}$ increases when the population increases.

Table 4. Comparison between the estimators of the total: adaptive sampling, SRS, adaptive SRS (ROOK)

Matrix	N	$N\bar{\mu}$	$N^2\widehat{var}(\bar{\mu})$	$N\bar{y}$	$N^2\widehat{var}(\bar{y})$	$N\bar{y}_{AD}$	$N^2\widehat{var}(\bar{y}_{AD})$
4x4	16	45.20	345.46	121.60	2,637.23	190.00	0
5x5	25	77.02	2,143.20	290.00	26,460.00	250.00	3,786.84
6x6	36	125.28	7,765.19	453.60	81,207.36	297.40	8,605.68
7x7	49	262.97	27,775.32	161.70	9,344.79	291.96	13,224.14
8x8	64	217.60	19,232.25	256.00	30,796.80	434.29	17,532.08
9x9	81	301.72	35,516.02	842.40	313,391.16	480.94	25,485.92
10x10	100	358.33	51,362.50	80.00	2,760.00	476.00	40,096.05
11x11	121	330.16	44,136.80	653.40	198,241.56	558.16	36,596.89
12x12	144	429.60	79,976.56	259.20	29,501.44	664.25	64,971.73
13x13	169	461.13	88,303.16	1,521.00	1,053,343.20	805.57	128,886.39
14x14	196	417.20	89,039.35	372.40	74,896.83	800.80	70,609.67
15x15	225	613.93	162,971.10	1,080.0	548,035.00	946.32	89,533.50
16x16	256	534.75	125,050.12	1,766.40	1,486,863.40	938.67	111,878.68
17x17	289	517.31	120,309.52	1,445.00	1,299,055.00	1,007.97	118,269.46
18x18	324	579.96	151,800.29	421.20	76,980.24	1,103.14	111,112.35
19x19	361	742.75	265,244.93	1,444.00	1,002,424.80	1,358.50	157,809.51
20x20	400	629.09	183,534.21	480.00	186,506.67	1,271.11	160,687.90

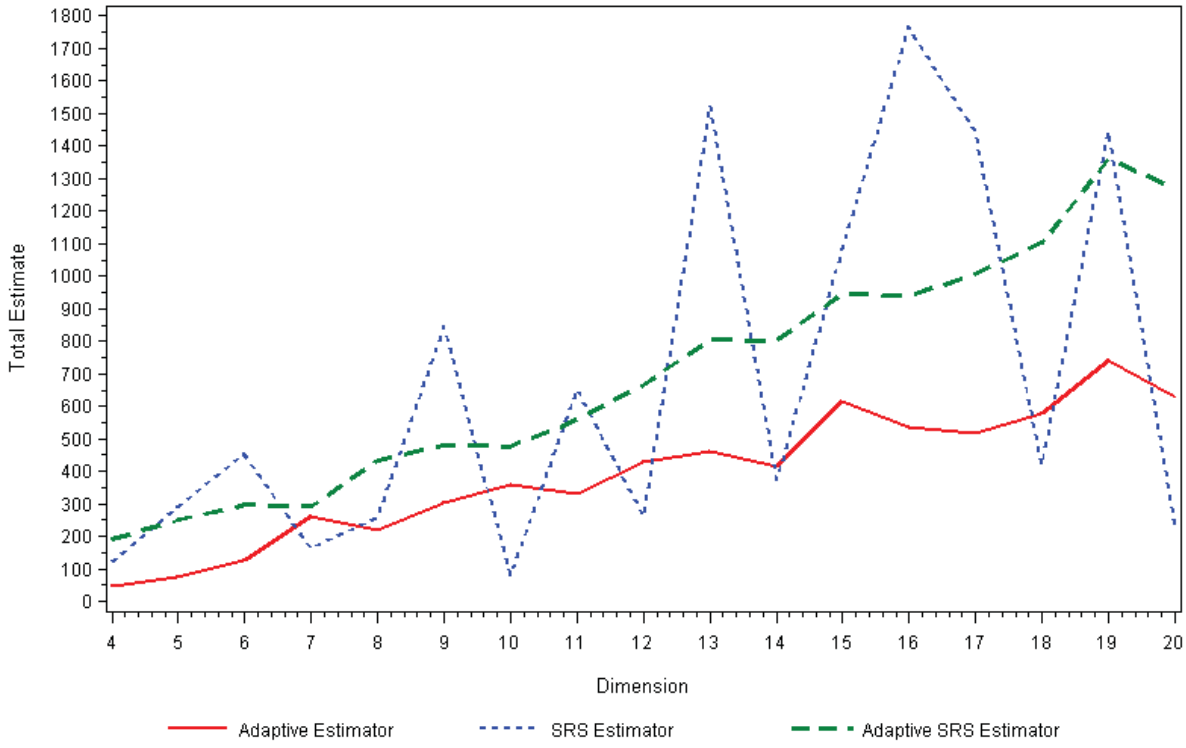


Figure 11. Analysis of the total when the population size increases (ROOK).

Figure 12 represents how the estimator of the variance in the estimator of the total changes with the variation in the population. We can observed that the estimator $N^2\widehat{var}(\bar{\mu})$ and $N^2\widehat{var}(\bar{y}_{AD})$ are closer and $N^2\widehat{var}(\bar{y})$ has large variation throughout the process.

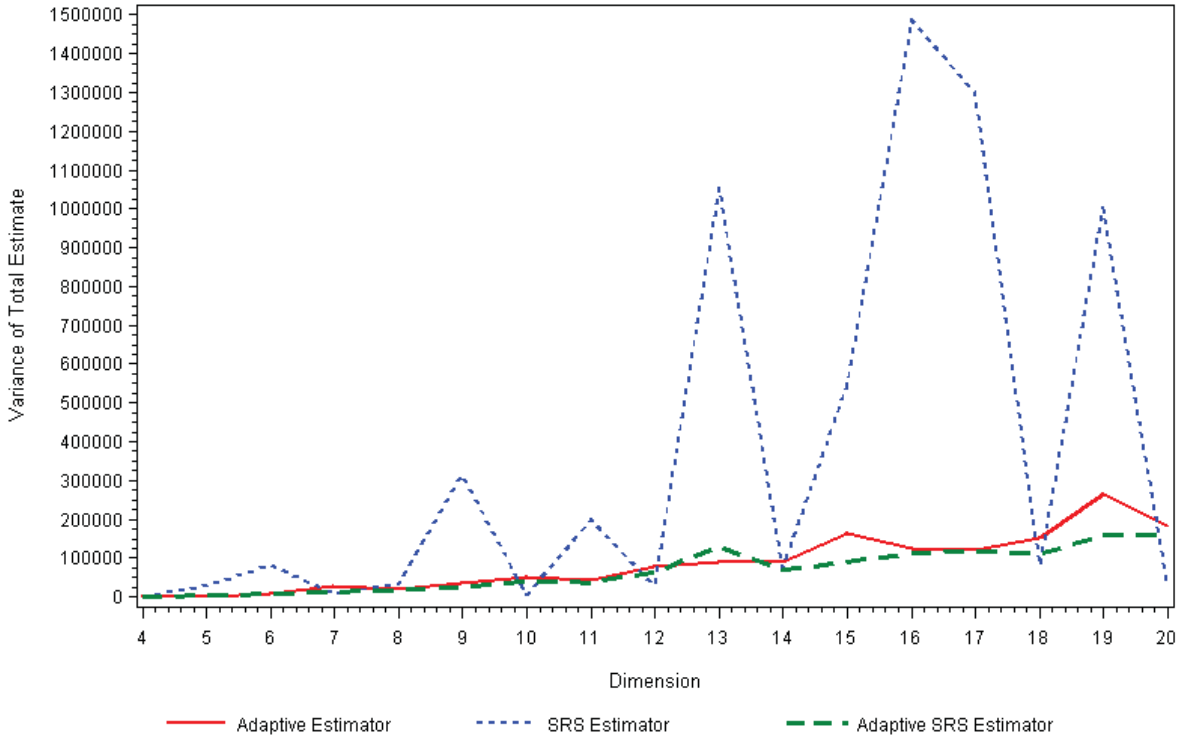


Figure 12. Analysis of the variance of the total when the population size increases (ROOK).

Figure 13 represents how the estimator of the variance in the estimator of the total changes with the variation in the population, with the largest values of $N^2\widehat{var}(\bar{y})$ removed. Notably, the variance is not constant as in Figure 12, showing disorganized growth for SRS and similar growth in the other two cases.

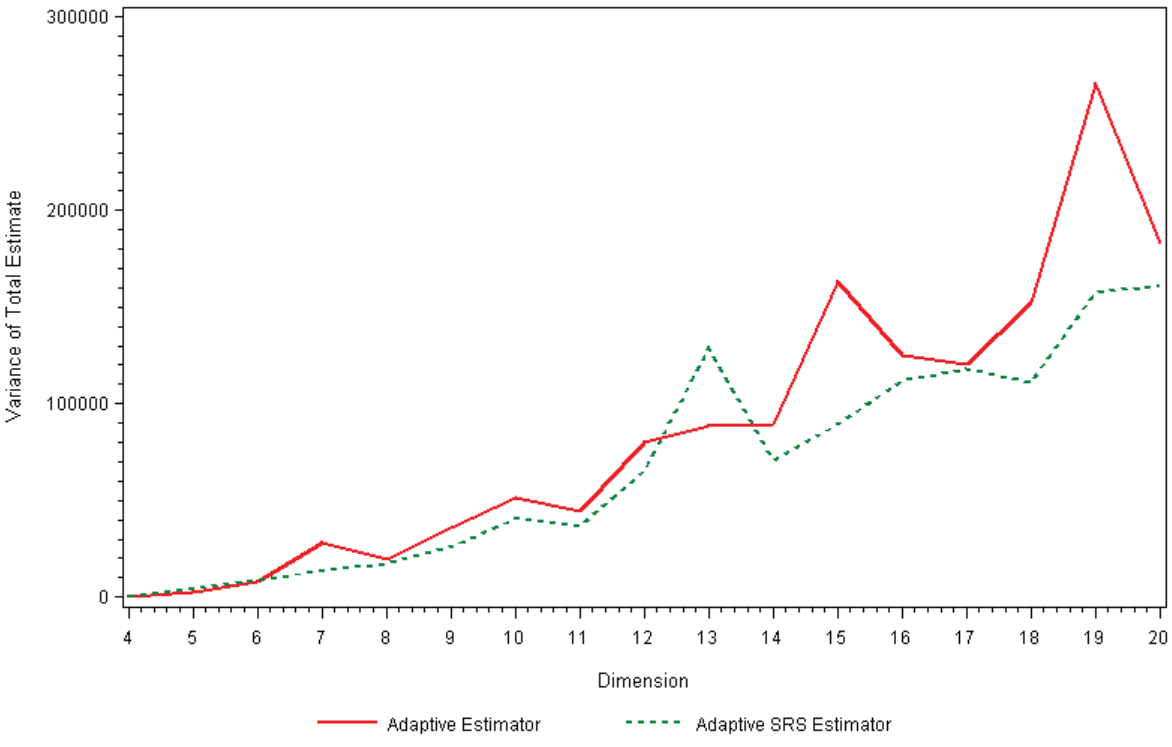


Figure 13. Analysis of the variance of the total when the population size increases (ROOK).

5.3 Stratified Adaptive Spatial Cluster Sampling

Thompson (1990) shows an example of how the stratified adaptive spatial cluster sampling technique works and compares the obtained results when considering whether the boundaries are present between the stratum. Initially, a regular grid is drawn on top of the area to be surveyed, and then, n units (squares) are selected by the SRS method.

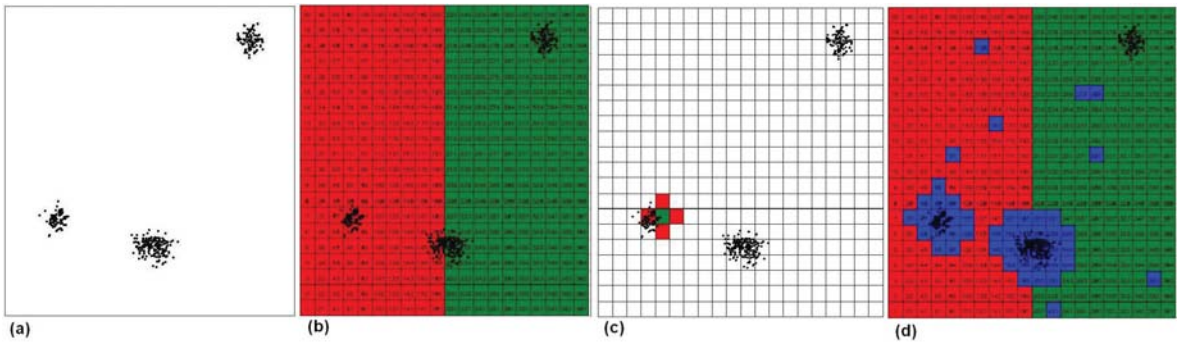


Figure 13. Stratified adaptive spatial cluster sampling.

The number of objects found in the analyzed area in Figure 14 (a) is 397 elements, within a total of $N = 400$ squares. Thus, it follows that the population mean is $\mu = \frac{397}{400} = 0.9925$. For this example, the region was divided into two strata, where for SRS, an initial $n = 10$ elements was selected with stratified adaptive spatial cluster sampling with equal sizes in each stratum. In stratum 1, we see a total of $N = 200$ squares and $n = 5$ elements for the initial sampling units, whereas in stratum 2, the others are $N = 200$ squares and $n = 5$, the total for the area.

As an example of adaptive spatial cluster sampling, the unit satisfies the condition if in each selected square one or more elements of interest is found. Because this condition is verified, it selects its neighbors. The neighborhood of each unit includes all adjacent units. Thus, a neighborhood can be analyzed in two ways: ignoring the existing boundary between strata to select the neighbors of a unit, or considering this limit.

In the first case, a unit to be selected as an element and that is in the square that has contact with the division of the stratum will have four neighbors - top, bottom, right and left - regardless of whether it is in a different stratum. Thus, the value of w'_{hi} for the estimator $\widetilde{\mu}'$, which ignores the limits of the strata, is zero for all units that do not satisfy the condition.

The first network intersect stratum 1, given in Figure 14 (d), and has a value of $w'_{11} = \frac{96}{6} = 16$. For the second network of intersection, the value is given by $w'_{12} = \frac{78}{5} = 15.6$, based only on the units of stratum 1. Thus, there is no intersection in stratum 2. Therefore, the estimate of the mean of the population and the estimated variance of $\widetilde{\mu}'$, given by Equations (35) and (37), respectively, is:

$$\widetilde{\mu}' = \frac{1}{400}\left[\frac{200}{5}(16 + 15.6 + 0 + 0 + 0) + \frac{200}{5}(0 + 0 + 0 + 0 + 0)\right] = 3.16$$

$$\widehat{var}(\widetilde{\mu}') = \frac{1}{400^2}\left[\frac{200(200-5)(74.9)}{5} + 0\right] = 3.65$$

where 74.9 is the variance of the five numbers $(16; 15.6; 0; 0; 0)$.

In the second case, a unit to be selected as an element and that is in the square that has contact with the division of the stratum will have three neighbors: top, bottom, right (or left), depending on whether this is in a different stratum. Thus, to calculate the estimator $\widetilde{\mu}$ (32), it is used for the same stratum $\frac{n_h}{N_h}$.

This obtains the variables w_{hi}, i.e., $w_{11} = \frac{96}{6} = 16$ for the first network and $w_{12} = \frac{192}{11} = 17.45$ for the second. The estimate for the mean and its variance given by Equations (32) and (33), respectively, is:

$$\widetilde{\mu} = \frac{1}{400}\left[\frac{200}{5}(16 + 17.45 + 0 + 0 + 0) + 0\right] = 3.35$$

$$\widehat{var}(\widetilde{\mu}) = \frac{1}{400^2}\left[\frac{200(200-5)(84.2)}{5} + 0\right] = 4.10$$

where 84.2 is the variance given by the five values of w_{1i}.

Table 5 presents the calculated estimates and Figure 15 shows the output of SAS; it appears that there is no significant difference between the estimators of the mean and variance of the estimated average if the boundaries of the strata are considered.

Table 5. Comparison between the estimators of stratified adaptive spatial cluster sampling.

Estimators	No crossing stratum boundaries	Crossing stratum boundaries
Estimators of the mean	3.16	3.35
Variance estimate of the mean	3.65	4.10

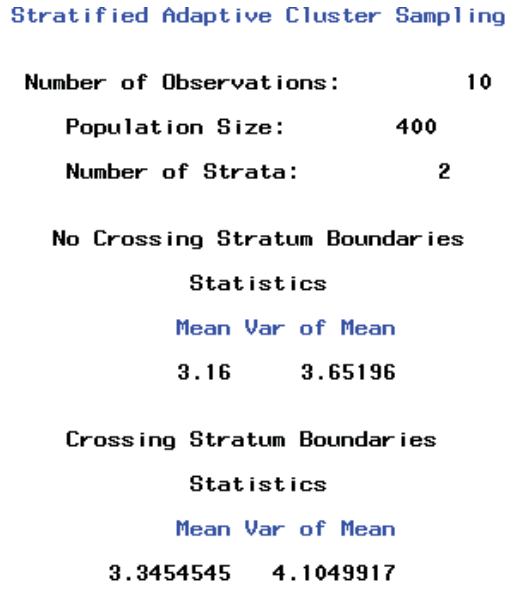


Figure 15. Output of the stratified adaptive spatial cluster sampling.

6. Final Remarks

This study shows that adaptive spatial cluster sampling suffers less variation between SRS and adaptive SRS, which demonstrates that it is a biased estimate. In Section 5, a comparison between different population sizes indicated that adaptive spatial sampling by cluster suffers less variation in the estimators than the other two samples.

Stratified adaptive spatial cluster sampling showed no significant difference between the estimators of the mean, or the estimated average, regardless of whether we consider the limits of the strata variances.

In conclusion, it follows that the computational algorithm for adaptive spatial sampling in this work is important, as this new technique has a variety of applications and users thus far do not have a computational tool to use it.

References

Brown, J. A. (1994). The application of adaptive cluster sampling to ecological studies. *Statistics in ecology and environmental monitoring, 2*, 86 C 97.

Brown, J. A. (1996). The relative effciency of adaptive cluster sampling for ecological surveys. Faculty of Infornia-tion and Mathematical Sciences.

Chang, M. (2008). Adaptive Design Theory and Implementation Using SAS and R. Chapman and Hall/CRC Biostatistics Series.

Chang, M. (2009). Adaptive design theory and implementation using SAS and R. CRC Press.

Cochran, W. G. (1977). Sampling Techniques (3rd ed.). Wiley.

Domingo, C., Gavald'a, R., & Watanabe, O. (2002). Adaptive sampling methods for scaling up knowledge discovery algorithms. *Discovery Science Lecture Notes in Computer Science, 6*(2), 131-152. http://dx.doi.org/10.1023/A:1014091514039.

Jain, A., & Chang, E. Y. (2004). Adaptive sampling for sensor networks. Proceedings of the first workshop on data management for sensor networks, 10-16.

Khan, A., & Muttlak, H. A. (2002). Adjusted two-stage adaptive cluster sampling. *Environmental and Ecological Statistics, 9*, 111-120. http://dx.doi.org/10.1023/A:1013723226430.

Kunigami, G. (2010, novembro). Ponto dentro de polígono. Technical report, Unicamp.

Ramsey, S. K. T. F. L., & Seber, G. A. F. (1992). An adaptive procedure for sampling animal populations. *International Biometric Society, 48*(4), 1195-1199.

Satyanarayana, A., & Davidson, I. (2005). A dynamic adaptive sampling algorithm (dasa) for real world applications: Finger print recognition and face recognition. *Foundations of Intelligent Systems (3488)*, 631-640. http://dx.doi.org/10.1007/11425274_65.

Seber, G. A. F. (1986). A review of estimating animal abundance. *International Biometric Society, 42*(2), 267-292.

Sengupta, R. N., & Sengupta, A. (2011). Some variants of adaptive sampling procedures and their applications. *Computational Statistics and Data Analysis, 55*, 3183-3196. http://dx.doi.org/10.1016/j.csda.2011.05.020.

Stein, A., & Ettema, C. (2003). An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agriculture, Ecosystems and Environment, 94*(1), 31-47. http://dx.doi.org/10.1016/S0167-8809(02)00013-0.

Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association, 85*(412), 1050-1059.

Thompson, S. K. (1991). Stratified adaptive cluster sampling. *Biometrika Trust, 78*(2), 389-397.

Thompson, S. K. (2011). Adaptive sampling. Technical report, Simon Fraser University.

Thompson, S. K., & Seber, G. A. F. (1996). Adaptive sampling. Wiley.

Waldispühl, J., & Ponty, Y. (2011). An unbiased adaptive sampling algorithm for the exploration of rna mutational landscapes under evolutionary pressure. *Research in Computational Molecular Biology Lecture Notes in Computer Science, 6577*, 501-515. http://dx.doi.org/10.1007/978-3-642-20036-6_45.

Yu, H., Jiao, Y., Su, Z., & Reid, K. (2012). Performance comparison of traditional sampling designs and adaptive sampling designs for fishery-independent surveys: A simulation study. *Fisheries Research, 113*(1), 173-181. http://dx.doi.org/10.1016/j.fishres.2011.10.009.

International Journal of Statistics and Probability; Vol. 4, No. 4; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Does the Constraint of Factor Loadings Impair Model Fit and Accuracy in Parameter Estimation?

Karl Schweizer[1], Xuezhu Ren[2], Tengfei Wang[1] & Florian Zeller[1]

[1] Department of Psychology, Goethe University Frankfurt, Frankfurt, Germany

[2] School of Eduation, Huazhong University of Science & Technology, Wuhan, China

Correspondence: Karl Schweizer, Department of Psychology, Goethe University Frankfurt, Frankfurt, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt am Main, Germany. Tel: 0049/69/798 35355. E-mail: K.Schweizer@psych.uni-frankfurt.de

Received: August 16, 2015 Accepted: September 2, 2015 Online Published: September 23, 2015

doi:10.5539/ijsp.v4n4p40 URL: http://dx.doi.org/10.5539/ijsp.v4n4p40

Abstract

The paper reports on the comparison of models of measurement with constrained and free factor loadings as part of confirmatory factor analysis in a simulation study. The comparison was conducted in order to find out whether constrained factor loadings that cause a reduced degree of adaptability to specificities of data mean a disadvantage in comparison to factor loadings that are freely estimated. Furthermore, the way of conducting the link transformation, the sample size and the number of variables were varied. The simulated data were dichotomous and constructed to conform to one underlying source of responding. The investigation of model fit and accuracy in estimating factor loadings yielded similar results for constrained and free factor loadings in confirmatory factor analysis. Furthermore, there were effects due to the type of link transformation and sample size.

Keywords: binary data, confirmatory factor analysis, link function, probability-based covariances, simulation, tetrachoric correlations

1. Introduction

The item discriminability is a characteristic of the model of measurement and reflects the relationship between the item and the corresponding latent attribute (Lucke, 2005). The model of measurement can show free or constrained discriminability. The shift from free discriminability to constrained discriminability reduces the adaptability to the specificities of data. Constrained discriminability characterizes the Rasch (1960) model and the corresponding one-parameter model (Birnbaum, 1968), the Rasch model-based linear logistic test model (Scheiblechner, 1972) and the tau-equivalent model (Lord & Novick, 1968). Despite the lack of discriminability, the Rasch model and the corresponding one-parameter model have so far played a major role in research and application guided by item-response theory (IRT), while the consideration of the linear logistic test model has been more or less restricted to investigations of specific effects, as for example the effects of the item position, learning and fatigue (Kubinger, 2008). These models are contrasted by the tau-equivalent model of measurement that can be employed for investigations in the framework of factor analysis. This model appears to originate from the relaxation of some constraints of the stricter parallel model of measurement, and the further relaxation of constraints leads to the less strict congeneric model of measurement (Graham, 2006).

It is the tau-equivalent model of measurement that is in the focus of the present investigation. The extended version of this model relates the $p\times1$ vector of observations $\mathbf{y}$ to the $p\times1$ vector of intercepts $\boldsymbol{\mu}$, the product of the $p\times q$ matrix of factor loadings $\boldsymbol{\Lambda}$ and the $q\times1$ vector of latent variables (i.e., latent factors) $\boldsymbol{\eta}$ and the $p\times1$ vector of error components $\boldsymbol{\varepsilon}$:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \ . \quad (1)$$

Since the tau-equivalent model of measurement is usually considered with respect to data that give rise to the expectation of one underlying source of responding, $\boldsymbol{\Lambda}$ is replaced by the $p\times1$ vector $\boldsymbol{\lambda}$:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \ . \quad (2)$$

The special properties that distinguish this model of measurement from other models are equally sized factor loadings $\lambda_1, \ldots, \lambda_p$:

$$\lambda_1 = \ldots = \lambda_p \tag{3}$$

and error components $\varepsilon_1, \ldots, \varepsilon_p$ that may differ from each other.

The measurement model shows constrained discriminability because of the restriction represented by Equation 3. In an application it can be realized in two ways: on one hand, the variance of the latent variable is set equal to one while the factor loadings are estimated under the restriction that they show equal sizes; on the other hand, the factor loadings are set equal to one while the variance of the latent variable is estimated.

One reason for the wide disregard of the tau-equivalent model as compared to the other models that also apply to dichotomous data is the lack of a link transformation. This transformation establishes a relationship between observed and latent scores with respect to scale and distribution. Such a transformation is characteristic of the generalized linear model (McCullagh & Nelder, 1985; Nelder & Wedderburn, 1972; Skrondal & Rabe-Hesketh, 2004). The Rasch model includes such a transformation. It is accomplished by means of the logit as link function. The same function also applies to the linear logistic test model. Since the original tau-equivalent model does not comprise a link function, its application for investigating dichotomous data can be questioned (Schweizer, 2012).

The measurement model must be complemented by a way of conducting the link transformation in order to overcome the differences between the categorical and continuous scales and the associated distributions. The first way is characterized by the combination of the tau-equivalent model of measurement as core part of confirmatory factor analysis with tetrachoric correlations as input (Muthen, 1984, 1993). In this case the computation of tetrachoric correlations includes the shift from the categorical scale to the continuous scale and from the binomial distribution to the normal distribution (Pearson, 1900). It is accomplished by means of thresholds that refer to the normal distribution function. The second way is due to the opportunity to modify the model of measurement in such a way that a link transformation is conducted additionally (Schweizer, 2013; Schweizer & Reiss, 2014; Schweizer, Ren, & Wang, 2015). In this case probability-based covariances that are also known as a pre-stage reached in computing the Phi coefficient (McDonald & Ahlawat, 1974) are recommended as input to confirmatory factor analysis in order to achieve interval scale. Their use presupposes a sample size of at least 200. Furthermore, weights reflecting the distributional differences between the dichotomous and continuous scores are included into the measurement model. The weights w_i that are assigned to the main diagonal of the $p \times p$ diagonal matrix $\mathbf{W}$ moderate the relationships between the observed and latent variables in such a way that there is an adjustment of the distributional differences between the variables regarding the variances and covariances:

$$\mathbf{y} = \boldsymbol{\mu} + \left(\mathbf{W}\boldsymbol{\lambda}\right)\boldsymbol{\eta} + \boldsymbol{\varepsilon} \ . \tag{4}$$

The weights transform the tau-equivalent model into a weighted tau-equivalent model.

However, even in the presence of an appropriate link transformation researchers and practitioners may suspect a low degree of efficiency and avoid using the tau-equivalent model of measurement because of its reduced adaptability to the specificities of data. There are a number of properties of data that potentially exert an influence on model fit (Tanaka, 1993), and these properties may be demanding to the adaptability of the model. Some properties appear to be especially important. First there is the correctness of the model with respect to the data at hand. The tau-equivalent model can be expected to perform well in correct models and to show a higher degree of sensitivity for incorrect models than the more popular congeneric model (Jöreskog, 1971).

Second the appropriateness of the construct representation that, in the absence of specific construct facets, pertains to the number of variables is important (Hogarty, Hines, Kromrey, Ferron, & Mumford, 2005; MacCallum, Widaman, Zhang, & Hong, 1999). To some degree the number of necessary variables depends on the response format. If the response format includes two categories only, usually a larger number of variables is necessary than in the multi-category case.

Third there is the sample size that is known to exert influence on model fit and is commonly varied in simulation studies (Bandalos & Gagné, 2012; Finney & DiStefano, 2013). The sample size must be large enough to achieve stable parameter estimates. The larger the size of the sample drawn from the population, the better is the chance to attain stable estimates. However, there is also a downside of a large sample size. The downside is the effect on the sensitivity of the chi-square statistic (Bergh, 2015) that plays a major role in the evaluation of model fit and contributes to several other fit statistics. In large samples the sensitivity of the chi-square statistic is usually so high that even a minor deviation of the model from data leads to the indication of model misfit.

In the following sections of this manuscript a simulation study is reported. It investigates whether constrained discriminability means an impairment of model fit and accuracy in parameter estimation (i.e., factor loadings) as compared to free discriminability if the data are dichotomous and the model is correct. In order to identify effects

of potentially moderating factors, the type of the link transformation, the sample size and the number of variables are additionally considered.

2. The Simulation Study

The major objective of the simulation study was to compare models of measurement with either constrained or free discriminability regarding model fit and accuracy in parameter estimation under the condition that the data are dichotomous and the model is correct. It required confirmatory factor analysis of simulated data on the basis of models of measurement with constrained and free factor loadings. The second objective was to investigate whether the result regarding the first objective depended on the type of link transformation. The possibility of a moderating effect of sample size on the result regarding the first objective constituted the third objective. Finally there was the objective to find out whether the number of variables servings as indicators to the latent variable exerted an influence on the comparison of constrained and free discriminability.

The four objectives defined the characteristics of the data that had to be generated. The first objective demanded that the data showed a one-dimensional structure that could potentially be identified by means of a one-factor model. Such a structure could be generated by means of a uniform pattern representing the relationships among random variables (Jöreskog & Sörbom, 2001, p. 159). According to the first and second objectives the data had to be dichotomous on one hand and to be related to continuous latent scores on the other hand. This combination of properties suggested the following theoretical background: originally continuous data did undergo dichotomization in the process of the assessment of a latent construct. Since dichotomization was likely to create different degrees of easiness or difficulty and since a broad range of easiness or difficulty was found to be quite demanding to data analysis including a link transformation (Schweizer, 2013), such a range was considered as a good precondition for the projected comparisons and investigations. For detecting a possible effect of the sample size, as was necessary according to the third objective, three different sample sizes were generated: $N = 300$, 1000 and 2000. Furthermore, the fourth objective required different numbers of variables for representing the latent variable. In accordance to the practices of scale construction six variables were considered to be close to the lower limit whereas twelve variables were accepted as good.

2.1 Data Generation and Analysis

Two uniform patterns establishing relationships among six or twelve variables respectively served the construction of the simulated data. These patterns were correlation matrices including the number .32 as off-diagonal elements. This number was expected to give rise to completely standardized factor loadings of .57 (Note. The expected size was .5656 that was selected because it could be expected to create a small degree of variability when rounding the estimated factor loadings. Such variability should prevent excessively high values in using Hartley's F_{max} test; for more details see below).

In the first step continuous random data following the normal distribution were generated: $X \sim N(0,1)$. The random data were arranged as matrices showing the following combinations of numbers of rows and columns: 300 x 12, 1000 x 12, 2000 x 12, 300 x 6, 1000 x 6 and 2000 x 6. Since 200 replications were considered sufficient for the cells of a design (Bandalos & Gagné, 2012), this number of matrices was generated for each combination. In the second step the continuous random data were recombined by means of weights computed for the two uniform patterns. These weights were achieved by means of a procedure described by Jöreskog and Sörbom (2001, p. 159). In the third step the continuous data were transformed into dichotomous data: $Y \sim Bin(2,p)$. In order to arrive at a broad range of easiness or difficulty, different proportions were selected for the columns of the matrices.

The transformation of continuous into dichotomous data in the third step was conducted according to the following proportions: .900, .827, .755, .682, .609, .536, .464, .391, .318, .245, .173 and .100 (i.e., number of selected events / number of all events) for matrices including twelve columns. In the first column of a matrix the 10 percent smaller numbers were transformed into zeros and the 90 percent larger numbers into ones. In the second column zeros replaced the 17.3 percent smaller numbers whereas the other numbers were changed into ones. The third to twelfth columns were processed in the same way in considering the other proportions of the list in corresponding order. In the matrices including six columns the proportions for splits were: .900, .740, .580, .420, .260 and .100. Subsequently, the dichotomous data served the computation of tetrachoric correlations and probability-based covariances. Additionally customary covariances were computed from the continuous data.

The objectives of the study required the consideration of four methods of conducting confirmatory factor analysis. These methods differed in characteristics of the model of measurement and the type of input matrix. The first method combined the congeneric model of measurement (Jöreskog, 1971) and tetrachoric correlations

as input (TetCon). The combination of the congeneric model comprising weights and probability-based covariances as input gave rise to the second method. The second model was addressed as weighted congeneric model (PbCWCon). Free factor loadings meaning free discriminability characterized the models of measurement of the first and second methods. The other methods differed from the described methods by using constrained factor loadings instead of free ones. Because of the assumed uniformity of the relationships among the variables induced by the uniform patterns, equal-sized constraints were necessary. The modification gave rise to the third method characterized by the constrained model of measurement and tetrachoric correlations as input (TetTau) and the fourth method that also included weighted and constrained factor loadings and probability-based covariances as input (PbCWTau).

Since all the methods included provisions regarding deviations from the normal distributions, the parameter estimation was conducted by means of the maximum likelihood method. The following fit indexes were considered in the evaluation of the results: chi-square, degrees of freedom, normed chi-square, RMSEA, SRMR, CFI, TLI and GFI. Cut-offs provided by Kline (2005) and Hu and Bentler (1999) served the evaluation of the results (RMSEA ≤ .06, SRMR ≤ .08, CFI ≥ .95, TLI ≥ .95, GFI ≥ .90). Because of the known dependence of chi-square on sample size two different cut-offs regarding normed chi-square were taken into consideration (2 for $N = 300$ and 5 for $N = 1000$ and larger). Means and standard deviations were computed for each set of 200 matrices. In order to facilitate the reading of the tables and the aggregation of the results, the superscript "M" was added to a mean if this mean proved to be good in comparison to the corresponding cut-off. If additionally the confidence interval meaning 95 percent of the distribution of the observed results was beyond the cut-off in the range of good results, it was replaced by the superscript "CI". For arriving at conclusions extending to different methods and data characteristics, the "M"s and "CI"s were counted in giving the weight 1.0 to each "M" and 2.0 to each "CI". Finally the counts were used for comparisons.

The accuracy of the completely standardized factor loadings was not only important regarding freely estimated factor loadings but also constrained ones because the average size of factor loadings varied as a result of parameter estimation and because of the item-specific contributions of error in standardization. Two aspects marked the investigation of the completely standardized factor loadings. First, the absence of the dependence on the item marginal (Kubinger, 2003; Torgerson, 1958) was checked. Dependence on the item marginal was obvious from a systematic deviation of the observed factor loadings from the expected factor loadings. In the simulated data equal sizes of the completely standardized factor loadings signified the absence of this kind of dependence. This check was conducted by means of Hartley's F_{max} test. Since the same size was expected for each completely standardized factor loading, the variances of these factor loadings obtained in investigating dichotomous data at the level of the means were compared with those achieved in investigating continuous data at the level of the means. In this check the F_{max} statistic served more as a descriptive statistic than a significance test. Since the variability of the means observed in continuous data that served as denominator of F_{max} was extremely small, it appeared to be overly sensitive to deviations. Second, the general size of the completely standardized factor loadings was checked.

2.2 The Results of Investigating Model-data Fit

In this section the results regarding model-data fit are presented. At first the results of investigating the covarian ce matrices computed from continuous data following the normal distribution by means of the congeneric model

Table 1. Means and Standard Deviations (in Parentheses) of the Fit Results for the Continuous Date in Different Sample Sizes (N_P = 300, 1000, 2000) and Numbers of Variables (N_V = 6, 12) Based on 200 Matrices

Input	N_V	N_P	χ^2	df	Normedχ^2	RMSEA	SRMR	CFI	TLI	GFI
Covariance	12	300	54.0 (10.3)	54	1.0^{CI}(0.2)	$.012^{CI}$(0.02)	$.033^{CI}$(0.00)	1.00^{CI}(0.00)	1.00^{CI}(0.01)	0.97^{CI}(0.01)
Covariance	12	1000	53.4 (10.9)	54	1.0^{CI}(0.2)	$.012^{CI}$(0.02)	$.018^{CI}$(0.00)	1.00^{CI}(0.00)	1.00^{CI}(0.00)	0.99^{CI}(0.00)
Covariance	12	2000	54.8 (10.9)	54	1.0^{CI}(0.2)	$.004^{CI}$(0.01)	$.013^{CI}$(0.00)	1.00^{CI}(0.00)	1.00^{CI}(0.00)	1.00^{CI}(0.00)
Covariance	6	300	9.1(4.8)	9	1.0^{CI}(0.5)	$.016^{CI}$(0.01)	$.025^{CI}$(0.01)	1.00^{CI}(0.01)	1.00^{CI}(0.02)	0.99^{CI}(0.01)
Covariance	6	1000	9.2 (4.5)	9	1.0^{CI}(0.5)	$.008^{CI}$(0.01)	$.014^{CI}$(0.01)	1.00^{CI}(0.00)	1.00^{CI}(0.00)	1.00^{CI}(0.00)
Covariance	6	2000	8.9 (4.0)	9	1.0^{CI}(0.4)	$.006^{CI}$(0.01)	$.010^{CI}$(0.00)	1.00^{CI}(0.00)	1.00^{CI}(0.00)	1.00^{CI}(0.00)

Note. ^{CI} The 95 percent confidence interval indicates good fit. ^{M} The mean indicates good fit.

of measurement are presented. They serve as a comparison level for the results obtained from dichotomous data, especially for gaining variances that are necessary for computing the F_{max} statistic.

The results presented in the upper half of Table 1 stem from matrices including 12 columns and the results of the lower half from matrices including 6 columns. The numbers that are not in parentheses are means and the numbers written in parentheses are standard deviations. All fit statistics obtained in investigating the continuous data and presented in this Table revealed that there was an overall good model-data fit. In all cases the confidence interval was completely below or above the corresponding cut-off. Furthermore, the sample size showed virtually no influence on model-data fit. Even the chi-square statistic did not display an influence of the sample size. Only the chi-squares observed in matrices including either 6 or 12 columns differed from each other. Matrices including twelve columns led to the larger chi-squares.

Table 2 provides the results observed in using the four methods described in the method section for investigating the matrices of dichotomous data comprising the larger number of columns ($N_V = 12$).

Table 2. Means and Standard Deviations (in Parentheses) of the Fit Results Obtained for Binary Data in Considering Different Parameters Types (PType), Link Types (LType) and Sample Sizes ($N_P = 300, 1000, 2000$) Based on 200 Matrices Including 12 Columns

PType	LType	N_P	χ^2	df	Normedχ^2	RMSEA	SRMR	CFI	TLI	GFI
Free[1]	TCorr[3]	300	343.3(198.2)	54	6.4 (3.7)	.124 (0.05)	.080[M](0.02)	0.77 (0.17)	0.72 (0.21)	0.85 (0.08)
		1000	418.4(293.5)	54	7.7 (5.4)	.078 (0.02)	.047[CI](0.01)	0.94 (0.05)	0.93 (0.06)	0.94[M](0.04)
		2000	346.3(163.6)	54	6.4 (3.0)	.051[M](0.01)	.031[CI](0.00)	0.98[CI](0.01)	0.97[M](0.02)	0.97[CI](0.01)
	WCov[4]	300	71.2 (12.5)	54	1.3[CI](0.2)	.016[CI](0.01)	.025[CI](0.00)	0.99[CI](0.02)	0.99[CI](0.02)	0.99[CI](0.01)
		1000	71.2 (13.8)	54	1.3[CI](0.3)	.016[CI](0.01)	.025[CI](0.00)	0.99[CI](0.01)	0.99[CI](0.01)	0.99[CI](0.01)
		2000	78.3 (14.6)	54	1.5[CI](0.3)	.014[CI](0.01)	.019[CI](0.00)	0.99[CI](0.00)	0.99[CI](0.00)	0.99[CI](0.00)
Fixed[2]	TCorr[3]	300	334.9(154.5)	65	5.2 (2.4)	.111 (0.04)	.097(0.02)	0.75 (0.18)	0.75 (0.18)	0.85 (0.06)
		1000	418.3(212.3)	65	6.4 (3.3)	.072 (0.02)	.058[CI](0.01)	0.94 (0.05)	0.93 (0.05)	0.94 (0.03)
		2000	369.1(144.6)	65	5.7 (2.2)	.048[M](0.01)	.039[CI](0.01)	0.98[CI](0.01)	0.98[CI](0.01)	0.97[CI](0.01)
	WCov[4]	300	72.5 (13.5)	65	1.1[CI](0.2)	.018[CI](0.01)	.057[CI](0.01)	0.98[M](0.02)	0.99[CI](0.02)	0.96[CI](0.01)
		1000	94.9 (15.1)	65	1.5[CI](0.2)	.021[CI](0.01)	.039[CI](0.00)	0.98[CI](0.01)	0.98[CI](0.01)	0.98[CI](0.00)
		2000	128.1 (17.6)	65	2.0[CI](0.3)	.022[CI](0.00)	.034[CI](0.00)	0.98[CI](0.00)	0.98[CI](0.00)	0.99[CI](0.00)

Note. [1] Free factor loadings. [2] Constrained factor loadings. [3] Tetrachoric correlations. [4] Weights and probability-based covariances. [CI] The 95 percent confidence interval indicates good fit. [M] The mean indicates good fit.

The first to sixth rows include results due to free factor loadings (TetCon, PbCWCon) and the seventh to twelfth rows contain results achieved by means of constrained factor loadings (TetTau, PbCWTau). The results of the second set of rows virtually mirrored the results of the first set of rows. The counting of the weighted "M"s and "CI"s of the first and second sets led to scores of 48 and 46 for free and constrained factor loadings respectively. These numbers indicated that there was virtually no effect due to the type of factor loadings. In contrast, the results that were specific for the types of link transformation suggested dissimilarity. The majority of the results based on tetrachoric correlations (TetCon, TetTau) reported in the first to third and seventh to ninth rows indicated a bad model-data fit whereas the link transformation by weights and probability-based correlations (PbCWCon, PbCWTau) reported in the fourth to sixth and tenth to twelfth rows signified good degrees of model fit. The corresponding overall scores were 23 and 71. There was also an effect of the sample size. The scores for the sample sizes of 300, 1000 and 2000 were 24, 29 and 41 respectively. According to these scores the increase in sample size improved the model fit.

Furthermore, matrices comprising the smaller number of columns ($N_V = 6$) were also investigated in order to find out whether the number of variables was important regarding model fit. The investigation of the matrices comprising 6 columns yielded the results of Table 3.

Table 3. Means and Standard Deviations (in Parentheses) of the Fit Results Obtained for Binary Data in Considering Different Parameters Types (PType), Link Types (LType) and Sample Sizes (N_P = 300, 1000, 2000) Based on 200 Matrices Including 6 Columns

PType	LType	N_P	χ^2	df	Normedχ^2	RMSEA	SRMR	CFI	TLI	GFI
Free[1]	TCorr[3]	300	129.1(101.2)	9	14.3(11.2)	.192 (0.08)	.115^{M}(0.07)	0.80 (0.13)	0.66 (0.22)	0.88 (0.08)
		1000	83.6 (91.1)	9	9.3 (10.1)	.082 (0.04)	.040^{CI}(0.02)	0.96 (0.04)	0.93 (0.07)	0.97^{M}(0.03)
		2000	75.2 (48.1)	9	8.4 (5.3)	.057^{M}(0.02)	.027^{CI}(0.01)	0.98^{CI}(0.01)	0.96^{M}(0.02)	0.99^{CI}(0.01)
	WCov[4]	300	9.4 (4.6)	9	1.0^{CI}(0.5)	.017^{CI}(0.02)	.032^{CI}(0.01)	0.98^{CI}(0.02)	1.00^{CI}(0.03)	0.99^{CI}(0.01)
		1000	11.5 (5.5)	9	1.3^{CI}(0.6)	.014^{CI}(0.01)	.019^{CI}(0.00)	0.99^{CI}(0.01)	0.99^{CI}(0.01)	1.00^{CI}(0.00)
		2000	14.2 (6.9)	9	1.5^{CI}(0.8)	.015^{CI}(0.01)	.015^{CI}(0.00)	0.99^{CI}(0.00)	0.99^{CI}(0.01)	1.00^{CI}(0.00)
Fixed[2]	TCorr[3]	300	187.5(138.0)	14	13.4 (9.9)	.185 (0.09)	.132 (0.05)	0.59 (0.32)	0.53 (0.44)	0.84 (0.10)
		1000	135.6(218.1)	14	9.7 (15.6)	.079 (0.05)	.061^{M}(0.03)	0.93 (0.11)	0.93 (0.11)	0.96^{M}(0.05)
		2000	93.6 (50.6)	14	6.7 (3.6)	.051^{M}(0.02)	.040^{CI}(0.01)	0.97^{CI}(0.02)	0.97^{CI}(0.02)	0.98^{CI}(0.01)
	WCov[4]	300	17.4(6.2)	14	1.2^{CI}(0.4)	.025^{CI}(0.02)	.053^{CI}(0.01)	0.96^{M}(0.04)	0.97^{M}(0.06)	0.98^{CI}(0.01)
		1000	25.3(7.2)	14	1.8^{CI}(0.5)	.027^{CI}(0.01)	.037^{CI}(0.01)	0.97^{CI}(0.02)	0.97^{CI}(0.02)	0.99^{CI}(0.00)
		2000	39.7(10.6)	14	2.8^{CI}(0.8)	.029^{CI}(0.01)	.033^{CI}(0.00)	0.97^{CI}(0.01)	0.96^{CI}(0.01)	0.99^{CI}(0.00)

Note. [1] Free factor loadings. [2] Constrained factor loadings. [3] Tetrachoric correlations. [4] Weights and probability-based covariances. [CI] The 95 percent confidence interval indicates good fit. [M] The mean indicates good fit.

This Table shows the same structure as Table 2. The results reported in this Table proved to be very similar to the results reported in Table 2. For example, the overall score of Table 2 was 94 and of this Table 93. Only the chi-squares included in the two Tables differed from each other considerably. However, this kind of difference was not unexpected and not considered as a problem regarding model fit.

Because of the high degree of similarity of results of the two Tables the final evaluation extended to the results of both Tables. First the scores characterizing free and constrained factor loadings (96 and 91) were compared by means of the chi-square test. The difference was not substantial (χ^2 = 0.13, df = 1, n.s.). Second the two ways of conducting the link transformation that yielded scores of 141 and 46 were compared and found to differ from each other (χ^2 = 48.26, df = 1, $p < .05$). Third there were the three sample sizes giving rise to the scores of 47, 58 and 82. The chi-square test indicated a significant difference (χ^2 = 10.27, df = 2, $p < .05$). Fourth the scores characterizing the different numbers of variables were compared. They did not differ from each other (χ^2 = 0.005, df = 1, n.s.). Because of the substantial differences regarding the ways of conducting the link transformation and the sample sizes the two types of factor loadings were also compared separately for each way of conducting the link transformation and for each sample size. In none of these comparisons a significant difference was observed (regarding the ways of conducting the link transformation: p = .77, p = .80; different sample sizes: p = .47, p = .79, p =.82).

In sum, constrained and free factor loadings as the two types of discriminability in confirmatory factor analysis do not differ from each other regarding model fit if the model is correct. Furthermore, the way of conducting the link transformation and the sample size influence model fit whereas the number of variables does not.

2.3 The Results of Investigating Accuracy in Parameter Estimation

In this section the results of investigating the accuracy of the completely standardized factor loadings are reported. The means and standard deviations of the completely standardized factor loadings achieved in investigating the covariance matrices computed from continuous data are presented in Table 4.

Table 4. Means and Standard Deviations (in Parentheses) of the Completely Standardized Factor Loadings Obtained for the Covariance Matrices (Mat) Computed from Continuous Data in Different Sample Sizes (N_P = 300, 1000, 2000) and Matrices Including 6 or 12 Columns Based on 200 Matrices

Matrix type	N_P	Position of factor loading											
		1	2	3	4	5	6	7	8	9	10	11	12
Covariance	300	.56(.05)	.56(.04)	.57(.05)	.57(.04)	.56(.05)	.56(.04)	.56(.05)	.57(.04)	.57(.05)	.56(.04)	.57(.05)	.56(.04)
Covariance	1000	.57(.02)	.56(.02)	.57(.02)	.56(.03)	.57(.03)	.57(.02)	.56(.02)	.56(.03)	.57(.02)	.56(.03)	.57(.02)	.56(.02)
Covariance	2000	.56(.02)	.57(.02)	.56(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)
Covariance	300	.56(.05)	.55(.05)	.57(.05)	.57(.05)	.56(.05)	.56(.05)						
Covariance	1000	.57(.03)	.56(.03)	.57(.03)	.57(.03)	.57(.03)	.57(.03)						
Covariance	2000	.56(.02)	.56(.02)	.56(.02)	.57(.02)	.57(.02)	.57(.02)						

The means varied between .55 and .57. This kind of variation was expected because of the necessity of rounding parameter estimates. The sample size did not show any influence on the variability of the means nor did the number of variables. However, a decrease of the standard deviation of the individual results from 0.05 (*N = 300*) to 0.02 (*N = 2000*) was observed when the sample size increased.

The results obtained in investigating the completely standardized factor loadings computed for the matrices of binary data including 12 columns are presented in Table 5. The first to third rows of this Table provide the completely standardized factor loadings for the tetrachoric correlations in combination with the congeneric model of measurement (TetCon). In the sample size of 300 the mean factor loadings varied between .49 and .57, in 1000 between .56 and .59 and in 2000 between .56 and .57. Furthermore, the F_{max} test results indicated deviations from the expected equality of the completely standardized factor loadings for the smaller sample sizes (*N = 300*: $F_{max}(2,11) = 167.37$, $p < .05$; *N = 1000*: $F_{max}(2,11) = 22.79$, $p < .05$; *N = 2000*: $F_{max}(2,11) = 1.15$, n.s.).

Quite different results were observed for the tetrachoric correlations in combination with the model including constrained factor loadings (TetTau) (see the fourth to sixth rows of Table 5). The mean factor loadings varied between .53 and .54 when the sample size was 300. In the other sample sizes the mean factor loading was always .57. The F_{max} statistic indicated a large deviation from the expected equality only for the sample size of 300 (*N = 300*: $F_{max}(2,11) = 133.37$, $p < .05$; *N = 1000*: $F_{max}(2,11) = 0.00$, n.s.; *N = 2000*: $F_{max}(2,11) = 0.00$, n.s.).

The completely standardized factor loadings obtained on the basis of probability-based covariances by means of the congeneric model (PbCWCon) are reported in the seventh to ninth rows of Table 5. In the sample sizes of 300 and 1000 the means varied between .56 and .58 and in the sample size of 2000 between .57 and .58. Investigating the dependence on the item marginal revealed significant deviations (*N = 300*: $F_{max}(2,11) = 9.06$, $p < .05$; *N = 1000*: $F_{max}(2,11) = 5.13$, $p < .05$; *N = 2000*: $F_{max}(2,11) = 3.71$, $p = .05$). In the case of the sample size of 2000 the ratio was only marginally significant, and in the other cases the F_{max} values were considerably smaller than the F_{max} values found for tetrachoric correlations.

Table 5. Means and Standard Deviations (in Parentheses) of the Completely Standardized Factor Loadings Computed from Binary Data in Different Sample Sizes (N_P= 300, 1000, 2000) and Matrices Including 12 Columns Based on 200 Matrices in Considering Different Parameters Types (PType) and Link Types (LType)

PType/	N_P	Position of factor loading											
LType		1	2	3	4	5	6	7	8	9	10	11	12
Free[1] /	300	.57(.16)	.53(.15)	.51(.12)	.50(.12)	.50(.11)	.49(.10)	.50(.11)	.49(.11)	.49(.11)	.50(.12)	.54(.14)	.56(.16)
TCorr[3]	1000	.59(.08)	.56(.05)	.57(.05)	.56(.04)	.56(.04)	.56(.04)	.56(.04)	.56(.04)	.56(.04)	.57(.05)	.56(.05)	.59(.08)
	2000	.57(.04)	.57(.04)	.56(.03)	.56(.03)	.56(.03)	.57(.03)	.56(.03)	.57(.03)	.57(.03)	.57(.03)	.57(.03)	.57(.04)
Fixed[2]/	300	.54(.08)	.53(.08)	.53(.08)	.53(.08)	.53(.08)	.53(.07)	.53(.07)	.53(.07)	.53(.07)	.53(.07)	.53(.07)	.53(.07)
TCorr[3]	1000	.57(.05)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.04)	.57(.05)
	2000	.57(.02)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.01)	.57(.02)
Free[1] /	300	.56(.09)	.57(.08)	.57(.07)	.58(.06)	.58(.07)	.58(.06)	.58(.07)	.58(.07)	.57(.07)	.57(.07)	.58(.08)	.57(.09)
WCov[4]	1000	.57(.05)	.56(.04)	.57(.04)	.58(.03)	.58(.03)	.58(.04)	.58(.04)	.58(.03)	.58(.04)	.58(.04)	.57(.04)	.57(.04)
	2000	.57(.03)	.57(.03)	.57(.03)	.57(.03)	.58(.02)	.58(.03)	.58(.02)	.58(.03)	.57(.03)	.57(.03)	.57(.03)	.57(.03)
Fixed[2]/	300	.57(.02)	.58(.02)	.58(.02)	.59(.02)	.59(.02)	.59(.02)	.59(.02)	.59(.02)	.59(.02)	.58(.02)	.58(.02)	.57(.02)
WCov[4]	1000	.57(.01)	.58(.01)	.58(.01)	.59(.01)	.59(.01)	.59(.01)	.59(.01)	.59(.01)	.59(.01)	.58(.01)	.58(.01)	.57(.01)
	2000	.57(.01)	.58(.01)	.58(.01)	.58(.01)	.59(.01)	.59(.01)	.59(.01)	.59(.01)	.58(.01)	.58(.01)	.58(.01)	.57(.01)

Note. [1] Free factor loadings. [2] Constrained factor loadings. [3] Tetrachoric correlations. [4] Weights and probability-based covariances.

Finally, there were the results achieved in investigating probability-based covariances in considering the model with constrained factor loadings (PbCWTau) (see the tenth to twelfth rows of Table 5). There was variation of the means between .57 and .59 in all sample sizes. Furthermore, the investigations of the equality of the factor loadings by means of the F_{max} test revealed substantial differences for all sample sizes ($N = 300$: $F_{max}(2,11) = 7.69$, $p < .05$; $N = 1000$: $F_{max}(2,11) = 7.29$, $p < .05$; $N = 2000$: $F_{max}(2,11) = 7.19$, $p < .05$). Given the significance of the F_{max} values, again it needs to be highlighted that these F_{max} values were small in comparison to the values observed for tetrachoric correlations computed from the smaller datasets.

Next the degree of correspondence of the overall observed and expected sizes of the mean factor loadings was evaluated in considering the different sample sizes. The tetrachoric correlations in combination with the congeneric model of measurement (TetCon) led to the mean factor loadings of .52, .57 and .57 (*SD = 0.12, 0.05, 0.03*) for the sample sizes of 300, 1000 and 2000. The replacement of the congeneric model by the model with constrained factor loadings (TetTau) yielded mean factor loadings of .53, .57 and .57 (*SD = 0.07, 0.04, 0.01*) for the sample sizes of 300, 1000 and 2000. The investigation of probability-based covariances in considering the congeneric model (PbCWCon) yielded mean factor loadings of .57 (*SD = 0.07, 0.04, 0.03*). These covariances in combination with the model including constrained factor loadings (PbCWTau) led to mean results of .58 (*SD = 0.02, 0.01, 0.01*) for all the sample sizes.

The matrices with 6 columns were investigated in exactly the same way as the matrices with 12 columns. The results of these investigations are included in Table 6. The results of the first to third rows were obtained using tetrachoric correlations in combination with the congeneric model of measurement (TetCon). In the sample size of 300 the results are based on 72 percent of the 200 matrices only. The other matrices were eliminated since the completely standardized factor loadings were either larger than 1.0 or smaller than -1.0. In the remaining matrices the mean factor loadings varied between .52 and .61. In the sample size of 1000 the means were between .55 and .60 and in 2000 between .56 and .57. Furthermore, the F_{max} test results indicated deviations from the expected equality of the completely standardized factor loadings for the smaller sample sizes ($N = 300$: $F_{max}(2,5) = 221.71$, $p < .05$; $N = 1000$: $F_{max}(2,5) = 94.87$, $p < .05$; $N = 2000$: $F_{max}(2,5) = 3.38$, n.s.).

Table 6. Means and Standard Deviations (in Parentheses) of the Completely Standardized Factor Loadings Computed from Binary Data in Different Sample Sizes (N_P = 300, 1000, 2000) and Matrices Including 6 Columns Based on 200 Matrices in Considering Different Parameters Types (PType) and Link Types (LType)

PType/	N_P	Position of factor loading					
LType		1	2	3	4	5	6
Free[1] /	300[5]	.61(.21)	.53(.13)	.52(.13)	.52(.14)	.52(.14)	.59(.20)
TCorr[3]	1000	.60(.12)	.55(.07)	.55(.07)	.55(.07)	.55(.08)	.60(.10)
	2000	.57(.05)	.56(.04)	.56(.04)	.57(.04)	.57(.04)	.57(.06)
Fixed[2] /	300	.61(.09)	.58(.06)	.58(.06)	.58(.06)	.58(.06)	.60(.08)
TCorr[3]	1000	.58(.04)	.57(.03)	.57(.03)	.57(.03)	.57(.03)	.58(.03)
	2000	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)	.57(.02)
Free[1] /	300	.55(.11)	.56(.09)	.58(.08)	.58(.09)	.57(.09)	.56(.11)
WCov[4]	1000	.57(.06)	.57(.05)	.58(.04)	.58(.05)	.57(.05)	.57(.05)
	2000	.56(.03)	.56(.03)	.58(.03)	.58(.03)	.57(.03)	.57(.04)
Fixed[2] /	300	.55(.03)	.56(.03)	.56(.03)	.56(.03)	.56(.03)	.55(.03)
WCov[4]	1000	.55(.02)	.56(.02)	.57(.02)	.57(.02)	.56(.02)	.55(.02)
	2000	.55(.01)	.56(.01)	.56(.01)	.56(.01)	.56(.01)	.55(.01)

Note. [1] Free factor loadings. [2] Constrained factor loadings. [3] Tetrachoric correlations. [4] Weights and probability-based covariances. [5] Datasets with factor loading >1.0 or < -1.0 were excluded leading to a reduction by 28 percent.

The replacement of the congeneric model by the model with constrained factor loadings (TetTau) led to quite different results (see the second part of Table 6). The mean factor loadings varied between .58 and .61 when the sample size was 300 and between .57 and .58 when the sample size was 1000. In the other sample size the mean was always .57. It indicated a deviation from the expected equality for the smallest sample size only ($N = 300$:

$F_{max}(2,5) = 22.72$, $p < .05$; *N = 1000*: $F_{max}(2,5) = 1.11$, n.s.; *N = 2000*: $F_{max}(2,5) = 0.00$, n.s.). Furthermore, the lack of variance in the sample size of 2000 signified that there was no deviation from the equality in size.

The investigation of the completely standardized factor loadings achieved for probability-based covariances and the congeneric model (PbCWCon) led to the results reported in the third part of Table 6. In the sample sizes of 300 there was variation between .55 and .58, in the sample size of 1000 between .57 and .58 and in the sample size of 2000 between .56 and .58. The investigation of the dependence on the item marginal revealed two substantial deviations (*N = 300*: $F_{max}(2,5) = 63.13$, $p < .05$; *N = 1000*: $F_{max}(2,5) = 4.23$, n.s., *N = 2000*: $F_{max}(2,5) = 14.10$, $p < .05$).

The investigation of probability-based covariances in considering the model with constrained factor loading (PbCWTau) yielded the results reported in the last part of Table 6. There was variation of the mean factor loadings between .55 and .56 in sample sizes of 300 and 2000 and between .55 and .57 in the sample size of 1000. Furthermore, the investigation of the equality of the factor loadings by means of the F_{max} test revealed no substantial deviation from the expected equal sizes (*N* = 300: $F_{max}(2,5) = 3.45$, n.s.; *N* = 1000: $F_{max}(2,5) = 2,96$, n.s.; *N = 2000*: $F_{max}(2,5) = 2.80$, n.s.).

Finally, there were the results of investigating the correspondence of the observed and expected sizes across different sample sizes. The tetrachoric correlations in combination with the congeneric model of measurement (TetCon) led to the mean factor loadings of .55, .57 and .57 (*SD = 0.16, 0.09, 0.04*) for the sample sizes of 300, 1000 and 2000 respectively. The replacement of the congeneric model by the model with constrained factor loadings (TetTau) yielded mean factor loadings of .59, .57 and .57 (*SD = 0.07, 0.03, 0.02*) for the sample sizes of 300, 1000 and 2000 respectively. In probability-based covariances in combination with the congeneric model (PbCWCon) all mean factor loadings were .57 (*SD = 0.09, 0.05, 0.03*). These covariances in combination with the model with constrained factor loadings (PbCWTau) led to .56 (*SD = 0.03, 0.02, 0.01*) for all sample sizes.

In sum, according to the results presented in the Tables 5 and 6 the constrained factor loadings showed consistency instead of variability in seven out of twelve cases whereas the free factor loadings in only three out of twelve cases. In six out of twelve cases the link transformation by means of tetrachoric correlations led to consistent factor loadings; otherwise consistency was found in four out of twelve cases. Regarding sample size there was mainly variability in the smallest sample size and mostly consistency in the largest one.

Most of the mean factor loadings were quite close to the expected factor loadings. Only the mean factor loadings computed for the smallest sample size and on the basis of tetrachoric correlations showed major deviations from the expected value. Furthermore, it needs to be highlighted that there were considerable differences regarding variability. In most cases constrained factor loadings were associated with smaller standard deviations than free factor loadings. Furthermore, the standard deviations of factor loadings due to weights and probability-based covariances were smaller than those of factor loadings based on tetrachoric correlations.

3. Discussion

Regarding the main objective that is the comparison of constrained and free discriminability realized as constrained and free factor loadings in confirmatory factor analysis, the outcomes of the simulation study suggest no difference. Constrained and free discriminability do not differ according to the degrees of model fit, and the result of the investigation of the parameter estimates is also suggestive of similarity instead of difference. In a way these results are surprising since free factor loadings mean a high degree of adaptability to specificities of data whereas the other type of factor loadings does not (Graham, 2006). The reason for the good performance despite constrained discriminability is presumably the correctness of the model. In correct models the kind of discriminability does not count. The difference between high and low adaptability to specificities of data is presumably important if the model is only partly correct. Consequently, constrained discriminability is more than just an option that is useful for overcoming estimation problems, as has been suggested (Millsap, 2001). Furthermore, constrained discriminability already proved useful in research using models designed to represent precise structural expectations (Schweizer, 2006a, 2006b, 2008, 2009).

The comparison of the results observed for different sample sizes reveals differences between the two ways of conducting the link transformation. The effect of the sample size is minor when the transformation employs weights in combination with probability-based covariances. But it is large for the transformation by means of tetrachoric correlations. Presumably the use of robust maximum likelihood estimation (Bryant & Satorra, 2012; Satorra & Bentler, 1994) would have improved the results, as is recommended by Finney and DiStefano (2013). However, there is also reason for abstaining from robust estimation. The first reason is that the theoretical foundation of tetrachoric correlation is so conclusive that insufficiency is not expected. Another reason is provided by the results of the simulation study for the largest sample size that are really good.

Finally, there are the results of investigating the possible dependence on the item marginals (Kubinger, 2003; Torgerson, 1958) by means of the F_{max} test. These results are important since the absence of dependence on the item marginals is indicative of the appropriateness of the link transformation. Without a link transformation the factor loadings of very easy and very difficult items are likely to differ considerably from those of the other items and give rise to a large variance. On the other hand a large variance assigned to the numerator of the F_{max} test leads to a large F_{max} value that means significance. However, significance needs to be interpreted with caution since the variance of the denominator of the F_{max} test that is computed from continuous data is very small.

Acknowledgments

The authors are grateful to the editor and anonymous reviewers for valuable comments.

References

Bandalos, D. L., & Gagné, P. (2012). Simulation methods in structural equation modeling. In R. H. Rick (Ed.), *Handbook of structural equation modeling* (pp. 92-108). New York, NY: Guilford Press.

Bergh, D. (2015). Chi-squared test of fit and sample size - a comparison between a random sample approach and a chi-square value adjustment method. *Journal of Applied Measurement, 16*, 204–217.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Bryant, F.B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling, 19*, 373–398. http://dx.doi.org/10.1080/10705511.2012.687671

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 439–492). Charlotte, NC: Information Age Publishing, Inc.

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement, 66*, 930-944. http://dx.doi.org/10.1177/0013164406288165

Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement, 65*, 202–226. http://dx.doi.org/10.1177/0013164404267287

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. http://dx.doi.org/10.1080/10705519909540118

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109–133. http://dx.doi.org/10.1007/BF02291393

Jöreskog, K.G., & Sörbom, D. (2001). *Interactive LISREL: User's Guide*. Lincolnwood, IL: Scientific Software International Inc.

Jöreskog, K.G., & Sörbom, D. (2006). *LISREL 8.80*. Lincolnwood, IL: Scientific Software International Inc.

Kline, R.B. (2005). *Principles and Practice of Structural Equation Modeling*. 2nd ed. New York: Guliford Press.

Kubinger, K.D. (2003). On artificial results due to using factor analysis for dichotomous variables. *Psychology Science, 45*, 106–110.

Kubinger, K. D. (2008). On the revival of the Rasch model-based LLTM: From constructing tests using item generating rules to measuring item administration effects. *Psychology Science Quarterly, 50*, 311–327.

Lord, F. M. & Novick, M. R. (Eds.) (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lucke, J. F. (2005). The α and ω of congeneric test theory: an extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement, 29*, 65–81. http://dx.doi.org/10.1177/0146621604270882

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84–99. http://dx.doi.org/10.1037/1082-989X.4.1.84

McCullagh, P., & Nelder, J.A. (1985). *Generalized Linear Models*. London: Chapman and Hall.

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27,* 82–99. http://dx.doi.org/10.1111/j.2044-8317.1974.tb00530.x

Millsap, R. E. (2001). When trivial constraints are not trivial: the choice of uniqueness constraints in confirmatory factor analysis. *Structural Equation Modeling, 8*, 1–17. http://dx.doi.org/10.1207/S15328007SEM0801_1

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous variable indicators. *Psychometrika, 49*, 115–132. http://dx.doi.org/10.1007/BF02294210

Muthén, B. (1993). Goodness of fit with categorical and other nonnormal variables. In K. A. Bollen & J. S. Long, (Eds.), *Testing structural equation models* (pp. 205–234). Newbury Park, CA: Sage.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, A, 135*, 370–384. http://dx.doi.org/10.2307/2344614

Pearson, K. (1900). Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London Series A, 195*, 1–47. http://dx.doi.org/10.1098/rsta.1900.0022

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

Satorra, A., & Bentler, P.M. (1994). Corrections to the test statistics and standard errors on covariance structure analysis. In A. von Eye & C. C. Glogg (Eds.), *Latent variable analysis* (pp. 399–419). Thousand Oaks, CA: Sage.

Scheiblechner, H. (1972). Das Lernen und Lösen komplexer aufgaben [Learning and solving complex cognitive problems]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 19*, 476–505.

Schweizer, K. (2006a). The fixed-links model in combination with the polynomial function as tool for investigating choice reaction time data. *Structural Equation Modeling, 13*, 403–419. http://dx.doi.org/10.1207/s15328007sem1303_4

Schweizer, K. (2006b). The fixed-links model for investigating the effects of general and specific processes on intelligence. *Methodology, 2,* 149–160. http://dx.doi.org/10.1027/1614-2241.2.4.149

Schweizer, K. (2008). Investigating experimental effects within the framework of structural equation modeling: an example with effects on both error scores and reaction times. *Structural Equation Modeling, 15*, 327–345. http://dx.doi.org/10.1080/10705510801922621

Schweizer, K. (2009). Fixed-links models for investigating experimental effects combined with processing strategies in repeated measures designs: a cognitive task as example. *British Journal of Mathematical and Statistical Psychology, 62*, 217–232. http://dx.doi.org/10.1348/000711007X268558

Schweizer, K. (2012). A weighted version of the tau-equivalent model of measurement for items with ordered response categories. *International Journal of Statistics and Probability, 1*, 151–163. http://dx.doi.org/10.5539/ijsp.v1n2p151

Schweizer, K. (2013). A threshold-free approach to the study of the structure of binary data. *International Journal of Statistics and Probability, 2*, 67–75. http://dx.doi.org/10.5539/ijsp.v2n2p67

Schweizer, K., & Reiß, S. (2014). The structural validity of the FPI Neuroticism Scale revisited in the framework of the generalized linear model. *Psychological Test and Assessment Modeling, 56*, 320–335.

Schweizer, K., Ren, X., & Wang, T. (2015). A comparison of confirmatory factor analysis of binary data on the basis of tetrachoric correlations and of probability-based covariances: a simulation study. In R. E. Millsap, D. M. Bolt, L. A. van der Ark, & W.-C. Wang (Eds.), *Quantitative Psychology Research* (pp. 273–292). Heidelberg: Springer.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modelling: multilevel, longitudinal and structural equation models*. Boca Raton, Fl: Chapman & Hall/CRC.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *testing structural equation m*odels, (pp. 10−40). Thousand Oaks, CA: Sage.

Torgerson, W.S. (1958). *Theory and Method of Scaling*. New York: Wiley.

International Journal of Statistics and Probability; Vol. 4, No. 4; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Imaginary Number Probability in Bayesian-type Inference

Richard Douglas[1]

[1] Independent researcher.

Correspondence: Richard Douglas, Republic Tatarstan, the Russian Federation. Tel: 7-906-11-22-749. E-mail: get@langlane.com

Received: July 12, 2015 Accepted: August 3, 2015 Online Published: September 23, 2015

doi:10.5539/ijsp.v4n4p51 URL: http://dx.doi.org/10.5539/ijsp.v4n4p51

Abstract

Doubt, choice and probability.

Bayesian probability computation is the most significant approach in complex maths interesting for all logicians to understand. And its computation and reasoning set us new priorities in further attempts to develop a more human-type reasoning, where 'possible' and 'probable' scales are matched and sorted out on subjective basis.

We use Bayesian computation models, Finetti's principle of free observation, dynamic probability, complex number equations, and other formal-logical principles in order to base our own modeling and sub-branching.

We aim to understand relation of the computation frequency in probability inference and in imaginary probability computation. And how the Bayesian inference principle could be disturbed by the possibilities of artificial 'doubt' of imaginary probability. We try to define the common patterns of complex number behavior in probability modeling, and the modeling of such probability in i numbers, so we could say one day that the probability of having a cancer is 1.99 in for 100, and the hypothetical probability of it is none (0).

The same subjective manner same subjective manner of a culprit who prefers an idea, or an image over logic, undertaking it as a guidance for his actions; a magnificent specter of a writer, a diamond of an artist and all those things which lure them all to the same jail of a culprit - the split of decision.

Keywords: Bayesian inference, imaginary number probability, Bayes theorem, conditional probability, imaginary probability, concurrent probability, inference

1. Imaginary Number Probability

1.1 Abstract Understanding

We understand imaginary unit of i as of any possible data that isn't included in factual observation by Bayesian inference, while the i^2, the imaginary number, or the 'non-existent' number is opposite - it is a possibility that could have happened but never 'will' or 'would' have.

Bayesian inference may not stipulate whether the $iP(A|B)$ is possible in equation, therefore, we have to expand the probability course in terms of abstract conditioning in the $(A|B)$ probability in where certain occasion may bring out hypothetical conditions of X1,2,3,....n.

$$\text{Condition } 1 \rightarrow \text{Supplication} = \text{Value of Condition } 1$$

So the condition of $(A|B)$ would be evaluated by $iP(A|B)$ as a valid one, and not only from numerical standpoint, but from the 'imaginary probability', from the cognitive 'intuition' if we may say so.

1.2 The Hypothesis of Imaginary Inference

We understand Bayesian inference in the complexity of statistical circumspection in decimal numbers of data withdrawal, but not as a dynamic variation of matching between the:

- most probable
- least probable
- possibly probable options,

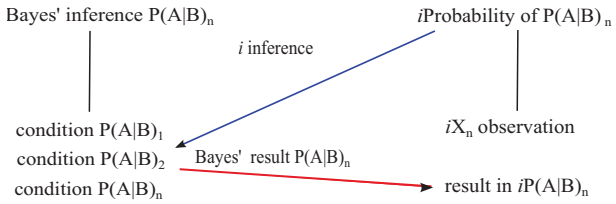


Figure 1

which are presumably superficial levels and would rather delay inference.

That's why we combine the principles of $P(A|B)$ inference into its possible 'imaginary' variances and 'imaginary' observations of i.

Hypothesis → proposition → observation → inference

Therefore, we have to simulate active inference-probability models on how a certain imaginary conclusion or value is possible in mathematical variation of computational logic.

1.3 Bayes' and de Finneti's Observations in Boolean Variances

According to the principles of De Finetti, we suppose factual changes in $Xi = 1$ by observation, (Finetti's exchangeability, more in Diaconis Persi (1977)), which comprises the premise for dynamic data observation in i number system.

And that is why we have to prove logically the free state observation of i number in $X_1 ... X_n$ in Bayes' conditional probability and applicably in $P(A|B)$ observation.

In order to gain the probability of assumption or the 'imaginary probability', simultaneously with the Bayesian inference process, we have to distinguish:

- precedent
- hypothesis
- type of data

See more on active observation in Robert F. Nau (2001).

2. Application of iP Probability in Complex Numbers

In order to predict the value, or the location, of any 'imaginary observation' we have to refer to the existing data, or to the existing precedent which we store in variances (v).

From the hypothesis to random observation we stipulate certain variances (v) into the frequency of them being hypothetical|probable and hypothetical|improbable in subjective inference. Hence, we stipulate an (iv) variance with the frequency (w) of logical requests (queries) in Bayesian $\frac{P(A|B)}{wi}$.

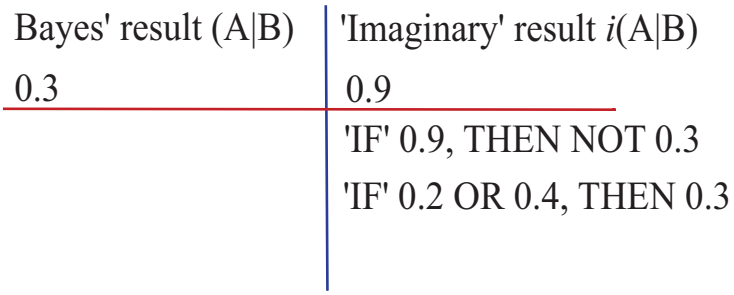
Respectively to the logic of computational differentiation there may be different values of inference: $X_n \subset Y_n \subset Z_n \neq 0$ in (x1, x2,.xn), and as for the calculus differentiations we may provide the f(x)=dx/dy differentiation in i; thus providing the 'unknown' integer into a $\prod$ of $P(A)_n$ or $P(B)_n$ in Boolean x,y data type.

We have to determine the velocity and the proportion of certain probability in iP in contra-pose to Bayesian analysis, in order to make conditional probability less systematic and more dynamic for random observation.

For example:

We have 3 doors and 2 of them are closed, what is the probability of the 3rd door being open?

We would say 1/3, or 0.3, or (10%), however it may be different from the cognitive standpoint. So, we have to presume reasonable factors over statistical, thence constructing stereotypical behavior in $iP(A|B)$:



*if the *i*P result is higher than P, then the foremost option is prefered

Figure 2

The factual probability that the door is open is 90%, 'because Terrence said so'.

Here occurs: "(B)=What if he lied?" and "(A)=We trust Terrence".

In either 'A' or 'B' options there would always be a variation of A1.2.3.4.5...n and B1.2.3.4.5...n.

We have to choose the $iP(A|B)$ counter-argument supplication of $P(A|B)$, in our example of $0.3i$ we stem out:

$$\frac{1}{3} = \frac{1(B)}{3(A)} = 0.3(10\%);$$

$$\frac{1}{3} = \frac{1(Bi)}{3(A)} = 0.9(90\%)$$

In where we take (A) as a factual info ('there are 3 doors'), and $(B)_n$ as a probable outcome.

Perhaps, we would understand such 'reverse' from B to $Bi(B \rightarrow Bi)$, and from 0.03(10%) to 0.9(90%) in the following supposition:

We suppose $(B)i$ as an x (a possible state of observation), and infer as $(B)i = xi_n$:

$$\frac{\frac{x}{3} \cdot 100}{3} = \frac{100x}{9}$$

We get roughly 9 actual variances in $100x$'s in observations. The further clarification may be related to the factual data observation of the factual $P(A|B)$, in order to get matched with the $iP(A|B)$.

For example, in $iP(A|B) \rightarrow P(A|B)$, regardless of its frequency (w), we may logically presume the closest value of its probability.

2.1 The i^2 Compromise of the i Value

Theorem 1 *The $i^2 = -1$ value while in $x \neq 0$, may be compromised - reversed.*

Proof. If we depict the i^2 in x tangent, we would depict it in both numerical values: positive and negative, proving that imaginary number exists in whole numbers, hence it's real.

While having negative value in i^2 $\sqrt{-}1$ $\sqrt{-}1$, we always yield minus, unless we specify it in:

$$(-\sqrt{x})i^2 = \sqrt{x}$$

In where we have to find the positive x out of the negative value:

$$\frac{\sqrt{x}}{\sqrt{x^2}} = \frac{1}{\sqrt{x}} \tag{1}$$

$$\frac{\sqrt{x^2}}{\sqrt{i^2}} = \sqrt{x^2} = x \tag{2}$$

$$\frac{\sqrt{-1}}{x^{-1}} = ix \tag{3}$$

Therefore, we would yield a supposition of $i^2 = x$, and $\frac{x}{i^2} = y$: (-1).

We may produce its results in (1) and (-1) for algebra and logic; in x (+), y(-) in trigonometry; and in P(A)=1, P(B)=-1 in Bayesian inference.

2.2 We Combine and Transfer i^2 Into i

If in $\frac{P(A|B)}{Xn}$ we set negative value collocation with positive ones in:

$$P(A|B)_v = \frac{i(-i)}{x} = \frac{1}{x} \tag{4}$$

$$X = \frac{P(A|B)}{P(A)} \tag{5}$$

$$iP(A|B) = \frac{i(i^2)}{x - i} \tag{6}$$

$$X = \frac{P(A|B)}{P(B)} \tag{7}$$

In where *iP* yields no productive result in *i*, therefore, needs to be regarded only hypothetically (least probable) or transferred into the *P*.

From here we try to achieve the similarity to the Bayesian probability model, only adding the *i* (sub)processing to it.

In $i = \sqrt{-1}$, we compose:

$$iP \to P = \sqrt{-1} = \frac{1}{x} \tag{8}$$

$$\frac{1}{x} = \frac{P(A|B)P(A)}{(x)} \tag{9}$$

$$iP = P(B_i) \tag{10}$$

$$P \to iP = \frac{P(A|B)(A)}{P(B_i)P(A)} \tag{11}$$

We gain such order of alignment of i^2 into *i* and vice versa. We do so in order to:

- avoid inapplicability of computation due to negative/non-existent value of it.
- have a substantiation of logical order.
- have comparison in the same system of $\mathbb{C}$ number (existing numbers).
- compare the inference difference in computational logic between $P(A|B)$ and $iP(A|B)$.
- manipulate further algebraic actions with the different constants, not only with $P(A|B)$, but with other models as well.
- sustain the role of i^2 in probabilistic models. Make it plausible, make it existent, to make it real and rational at the same time.
- and finally, we transfer a negative number of i^2 into the *i* number, which is imaginary but existent (!), hence compatible with Bayesian analysis computation.

2.3 Dynamic State Observation

We know that we're containing the layer of the i^2 into the computable value in dynamics, we may presume the negative value of *i* in $\mathbb{C}$ computation of $\sqrt{-1} = \frac{ix}{n}$, so it would be a platform for $\mathbb{C}$ equations, where we presume the *i* derivative of $(A)|(B)$, whereas (A)=1, (the factual probability) in opposite to (B)=-1 of nonexistent probability:

- (A)=1; in *iP*
- (B)=-1; in *iP*

in $\mathbb{C}$ computation:

$$i^2 = \frac{-1}{x} \to (B) \tag{12}$$

$$\frac{-1}{xi^2} = \frac{1}{x} \to (A) \tag{13}$$

Lemma 1 *If we presume an 'imaginary' dynamic state observation in* $\frac{i}{1x}$*, then we would yield an ix probabilistic reasoning in* $iP(A|B)$ *of Xn computation.*

Proof In $\frac{i(x)}{0.3} = \frac{i(x)}{3}$, where x is the state of observation, could be drawn out as in the following (the '3 doors' example):

$$\frac{(ix) \cdot (0.3)}{(100\%) \cdot (3)} = 9 \tag{14}$$

Hence, we proportion the bundle of the '3 doors' to the average index of 0.3 in proportion to the 100% of its credibility.

And thus, the negative value of i shifts the observation of x in $(ix \to Xn)$, where we have only 9 variances of imaginary probability instead of ∞. So, we bind it into stereotypical reasoning.

> WE HAVE 3 DOORS, ONE OF THEM IS 'X'. WE YIELD 9 PROBABILITIES ON WHY IT COULD BE OPEN, AND ONLY 1 ON WHY IT COULD BE SHUT.

3. Inference Programming

3.1 The Limitation of P(A|B) in P[X,Y,Z]

A practical solution in $i^2 \to i$ transfer in computation may lead us into tremendous suppositions of 'intuition thinking' and AI automated inferences.

Nevertheless, the intuition levels of Bayesian inference may be expressed in mathematical limitations of [X, Y, Z] variables, accorded by the mathematical value of x (random observation, also Finetti) in i^2 and i of $iP(P(A|B)$. Hence, may be principled in:

$$\sum_{i=1}^{N} X_i \to E[Xn], \quad N \to N_{x,y,z}. \tag{15}$$

Where the frequency (w) of iX may be reducible in $i(P(A|B)$ in order to make the latter more intuitively limited,

$$\sum_{iX=x(w)}^{N} X_i \to iX[ixw], \quad N \to N_{x,y,z}. \tag{16}$$

$$\sum_{i=1}^{N} X_i \to E[Xn], \quad N \to N_{x,y,z}. \tag{17}$$

Whereas in $iX = x(w)$ we may have [X,Y,Z] variables only, in where the reducibility of X_i is the reducibility of a variable, but not a state of observation x.

Another example of it exists in the P(Yn, X1) in Carbonari A. and Giretti A. (2014), in where the static Y is observed by the dynamic y, and the comprehension of $P(X|Y)$ simply goes by the $\sum$ of variable in $P(A|B)$.

However, even such distant observation is processed via the given data of $P(X|Y)$ in where only X and Y observe each over, and in where it is hard to grasp intuiting levels.

Eventually, we could limit ourselves to the existing conclusion of:

$$P \to \infty$$

But we may also have to understand it in:

$$P \to N_{x,y,z}.$$

And we'd rather proceed to the stochastic modeling in Bayesian inference computation, instead of calculating infinite maths in maths. So, we pre-set:

$$P(X|A)$$

$$P(Y|B)$$

$$P(Z|v)$$

3.2 The $iP(A|X)(B|Y)$ in Stochastic Computation

If we know that its inference reducible in Bayes' conclusion (as well as in HMM) and multiplied via free-state observation (Finetti) wouldn't that logically preclude that there is a certain probability (Bayes') and at the same time uncertain variation of it (Finneti), hence no probability at all?

There wouldn't, unless we specify the free variance v in a certain probability $P(A|B)$ of an uncertain data iPn/iPx in it.

It's presumed, as soon as, any physical observation is already bound into [X, Y, Z] variances of multiple observations of x.

Confusing or not, the goal is to establish the iP and P probabilities in decimal and whole number computations. And we still have to deal with negative value of $i = i - 1$ in it.

Is it anyhow possible to presume that P=-1? Perhaps in binomial equation of it:

$$\sum_{k=-1}^{n} = 1 \rightarrow \frac{P(A|X)_i(B|Y)}{P(B)} \tag{18}$$

$$\sum_{k}^{n} = -1 \rightarrow \frac{P(A|X)(B|Y)}{P(B)_i} \tag{19}$$

In the following we try to elaborate the computational $index_i$, whereas we set:

$$IF, P = i^2, THEN : P =' -1' \tag{20}$$

$$\frac{P(A|B)}{(A|B)_i} = -1, OR' 1' \tag{21}$$

$$IN(A) = 100, (B) = 40; (A_i) = 80, (B_i) = 20. \tag{22}$$

$$IF' 1', THEN : P(A|X) \rightarrow \frac{P(A)}{(B)} \tag{23}$$

$$P(A|X) = \frac{100}{40} = 2.5; \tag{24}$$

$$IF' 1', THEN : iP(A|X) \rightarrow \frac{P(A)(B_i)}{(B)} \tag{25}$$

$$\frac{(100) \cdot (80)}{40} = 2.0; \tag{26}$$

$$IF' - 1', THEN : P(A|X) \rightarrow \frac{(B)}{(A)} \tag{27}$$

$$\frac{40}{100} = 0.4; \tag{28}$$

$$IF' - 1', THEN : iP(A|X) \rightarrow \frac{(B)(A_i)}{(A)}; \tag{29}$$

$$\frac{(40) \cdot (80)}{100} = 3.2. \tag{30}$$

In where the portability of $P(A)^n$ varies both in 2.5/2.0i and in $P(A)^{n-1}$ in 0.4/3.2i; whereas we consider the probability of 3.2i for $iP(A)^n$ more likable than 2.5 and 2.0i, because 3.2i as close to 2.5 as possible (the 0.4 is the least probable).

3.3 Index Proof in $iP(A|X)(B|Y)$

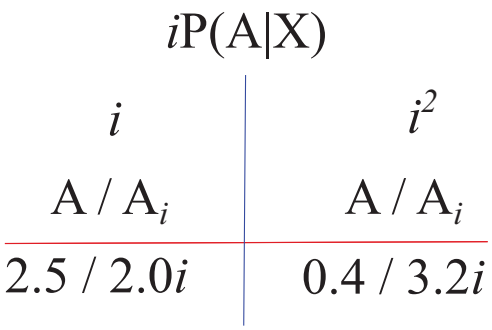


Figure 3

If we stipulate an average coefficient in A/A_i from the given data of $P(A)^n$ (2.5) and $iP(A)^{-1}$ in (3.2i) on the same matter:

$$\frac{P(A)(iP(A|X)}{iP}$$

in multiplication of:

$$\frac{2.5 \cdot 3.2.}{100(\%)} = 0.08$$

while in:

$$\frac{2.5 \cdot 2}{100(\%)} = 0.05$$

$$\frac{2.5 \cdot 0.4}{100(\%)} = 0.01$$

There is a chance of imaginary probability to take place over the factual one.

The similar applies to $(B)|(B_i)$ in:

$$\frac{P(B)(iP(B|Y)}{iP}$$

in where you may be given different data, but a similar decimal outcome.

Another step, is to initiate both probabilities in i and i^2 as in $P(A|B)^{n-1}$ and $P(A|B)^n$ of $(A|X)(B|Y)$ and formulate them in:

$$iP(A|X) = \frac{P(B)_i(P(A)}{P(B)}$$

$$iP(B|Y) = \frac{P(A)_i(P(B)}{P(A)}$$

and for the separate coefficients in A/A_i and B/B_i:

$$indexA/A_i = \frac{P(A)(iP(A|X)}{iP}$$

$$indexB/B_i = \frac{P(B)(iP(B|Y)}{iP}$$

3.4 Stochastic Reasoning and the i Index

If there is any data given to us by $P(A|B)$ computation would that rouse any inference in data non-existent, in cases of artificial doubt, inference or recollection?

We may always differentiate (A) on (B) and (B) on (A), but we wouldn't ever get as close as possible to the probabilistic reasoning, unless we require a mathematical product to do so:

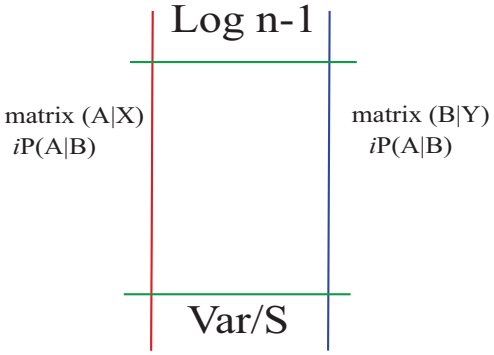


Figure 4

$$\prod_{k=1}^{n} = 1 \rightarrow P(A|X)(B|Y)_i \tag{31}$$

$$\prod_{k}^{n} = -1 \rightarrow iP\left(Y|i^2\right)(X|A) \tag{32}$$

Determining the two fields of probability $(X|A)(Y|B)$, we may use them in a more dynamic way:

$$\sum_{k=1}^{k} = 1 \rightarrow P : \frac{(A|X)(B|Y)}{P(B)v} \tag{33}$$

$$\sum_{k}^{k} = -1 \rightarrow P : \frac{(A|X)_i(B|Y)_i}{(A|X)v} \tag{34}$$

In Paisley J. and Jordan M. (2012) this concept reviewed as the stochastic search in programming.

The other practical and foreseeable possibilities of iP in computability are traced in Claret G. and Sriram K. (2013) models, in where the sequential Bayesian inference in programming is a subset of $\mathbb{R}$ numbers.

$$\frac{MatrixA}{MatrixB} \rightarrow Variation \rightarrow Observation \rightarrow Inference$$

3.5 Inference and the Frequency

In conditional probability modeling we may express the variance/var/(v) in mathematical frequency (w), so, that would yield progression in Log_{n-1}, the linear computation.

As we understand the linear inference, handled by the inference-algorithm, it's quicker than the random observation.

However, we may consider the following computational pacing similar to the real-time (w) progression of the free-bound variable in Bayesian type probability:

$$s_{n=1}^{n} \rightarrow \frac{P\,(A|X)\,(B|Y)}{wLog - 1} \tag{35}$$

$$s_{n=1}^{n} \rightarrow \frac{P\left(A|X^1\right)\left(B|Y^{-1}\right)}{P(B)_w Log x_n} \tag{36}$$

We may consider a time-out in real-time $(A|B)$ inference if only there is any limited data in $P(A|B)$ on the other hand; then the iP would be hypothetically infinite for computation, but it shall be confound in the spectrum of [X,Y,Z] instead.

$$s_{n=1}^{n} \rightarrow \int B\frac{n}{y}|A\frac{n}{x} \tag{37}$$

Composing of $x \in y, z$ and $y \in z$, we may further consider the frequency of Bayesian probability in its imaginary unit as following:

$$P(A|B) \to \frac{P(A|X)\left(B|Y^{-1}\right)P(A|Z)}{P(B|Y)Log-1} \tag{38}$$

3.6 Measure-theoretic in iP Programming

Concerning the time-frequency relation in negative value of i^2, expressed through binomial fraction either via integer frequency, we stipulate it as following:

$$\int \frac{dx^n}{Log_n} \tag{39}$$

$$S \to \int \frac{dx^2}{Log-1} \tag{40}$$

Theorem 2 *The frequency* (w) *reducible to the extent of negative number in* $P(|B)$ *conditioning, and yields a specific result of preference (choice) between probable results.*

Proof. We consider an iP inference in $\frac{iP(w)}{P(w)}$ frequency, and the $\frac{ix}{x}$ derivatives in Bayesian probability, in where we set $X(1)|Y(-1)|Z(log1)$ in order to yield iP in both results: in $x > 1$ and in $y < 1$, whereas $X \in Z \geq Y$.

From the existing model of $Pr\int[X \in A]$ we would try to construe its specifications in log value for our $P \to iP$ negation:

$$Pr\int[X \in A] = \int_X^{-1} P(B) = \int[log-1] = \int_A [A|B]log-1$$

Specifying that the $A|B \in i$, we would later on proceed at any computation of Bayesian inference.

3.7 Time Reducibility in iP

In the timing of the reducibility of variances, we understand the iP inference as imaginative, hence, existent only at particular time (t) and in a particular point of data of $A|$ or $B|$. In where we compile a nonlinear shift of (t)-time and (v)-variance in bound of $(A|B)$:

$$\int\int\int \frac{AvXn}{t \in A(ZY)}$$

3.8 Inference, Time, Probability

Timing recurrent, the solid statement is given, a mind is the inference, a blast is the movement of variances v^n:

$$\int\int\int \frac{AvXn^1}{(B|X)Yv}$$

The system of it simplified to the existent presence of science:

$$\Pi_n \int\int\int \frac{Av|X^{n1}}{(B|X)(B|Yv^n)(log-1)} \tag{41}$$

$$\Pi_n \int_v SX \to \frac{\mathbb{R} \to i}{v} \tag{42}$$

The $\mathbb{R} \to i$ reducibility is the reducibility of time. We contemplate not the data but the time. Therefore, it is a question of a philosophy to proceed further on, while us consider the logical order of it.

4. The Time Occurrence and the Variances

4.1 Getting an Average Index

We review variances (v), time (t) and the frequency (w) in derivatives of X_n of computational data, and bound variables of [X,Y,Z] in certain logical calculations, and we presume the probability in i in conditioning of time to it.

We structure our hypothesis in $P(A|B)$ reducibility to [X,Y,Z] in Bayesian network, time frequency and time reducibility to certain extent.

The imaginary probability variance considered to be a pooled variance of $P(A|X)(B|Y)$ that is to be matched further on by the CPU or by the time-out command.

If there is a certain imaginary probability occurs in processing as of the '3 door' example, there might be consideration of timing (tv) in:

$$\int_v SX \to \frac{\mathbb{R} \to i}{v}$$

so that the $\mathbb{N}$ number of the occurrence would ignite the actual value of probability.

Lemma 2 *For example, if we know that the probability of having cancer is 90% (A) and not having it 10% (B), then it would be only a generalized data, while having it in time occurrence may presume such data as subjective, 'connected' to the observer.*

Proof. An average index multiplied by occurrences per time may have a statistical probability for subjective preference in counterpoint to the factual data.

Then what occurrence of (A) may be allowed for the observer of computation? If 90% of it is the value of conclusion, then we have to consider the average time occurrence of such probability:

$$iPt = \frac{(A)}{tv}$$

Variances 3 types of cancer. We stipulate imaginary time occurrence in:

$$iPt = \frac{(A)}{tv} \tag{43}$$

$$iPt = \frac{A}{tv} = \frac{iPt(B)}{A} \tag{44}$$

From the given table we fit the data:

$$\frac{90}{2 \cdot 6} = 7.5 \tag{45}$$

(May get 75% of chances to 'occurring/not occurring' at the moment of contemplation in iPt (A).

$$\frac{7.5 \cdot 10}{90} = 8.3 \tag{46}$$

(May get as 83% of occurrence including the probability (B)) Finally we get 90% by Bayesian inference vs 83% of imaginary probability.

If we calculate probability by Bayes' principle, we still have to proceed on which data we choose for. There may be several results, and only one that fits best for the observer may be valid.

4.2 Linear and Non-linear Understanding of Time in iP

In both, linear and non-linear computations we set $Xi\ Yi$, so the variance of them would be in a proper time-set, whereas in "density function" explained in Noack B. and Klumpp V. (2008) we would yield:

$$P := \sum_{i=1}^{n} \alpha_i f_i | \sum_{i=1}^{n} \alpha_i = 1, \alpha_i \geq 0$$

In analogy to the model above, we chase parallel in:

$$P = \sum_{i=1}^{n} f_n \alpha_n | \sum_{i=1}^{n} f_i = \alpha_i \leq 0 \tag{47}$$

In where $f_i....f_n$ construe [X, Y, Z] derivatives pertinent to its iPt (time) only in α, whereas $\alpha = Y|(B)$, or any other $\mathbb{R}$ number.

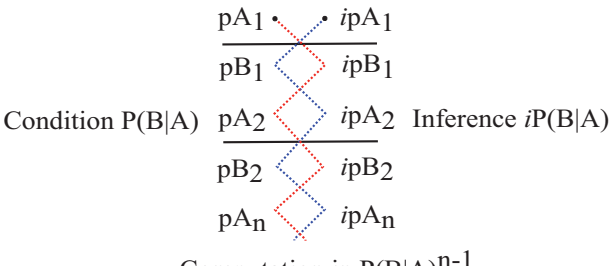


Figure 5

Therefore, we apply the principle of *XiYi* /per time, and define the (B) (e.g. bias) from $i^2 \rightarrow i$ and to natural numbers. Because in $\mathbb{C}$ numbers of any automatic equation it's the value of $\mathbb{C} = 0$ or $\mathbb{C} \neq 0$ that signifies the value probability.

$$P = \sum_{i=1}^{n} \int \frac{f_n}{\alpha_i} = \int \frac{f_n \alpha_i}{iPt} \tag{48}$$

Where basically proceed into relation of fa_i with iPt, which was explained previously in $\frac{iPt}{v}$.

4.3 Further i Integration into Bayesian Conditioning

The *i* number integration in $\mathbb{C}$ number derivatives in the following example of: 3.0 + 4.0i = 3.0 + 4.0 * Complex I [1], would specify the *i* as a part of computational integer in Bayesian logic, hence the adaptability to its programming in probability reasoning:

$$\frac{P(B|A)}{iP(B|A)}$$

We file 'Condition 1' of P(A) P(B) into imaginary inference of $iP(A)iP(B)$ in the spectral alignment (Fig. 5), then we parse the derivatives and proceed them onto separate computations.

Such schematic conditioning of binomial probability is mutually exclusive, and the conditional probability of $iP(B|A)$ would draw the nearest inference possible for its further computation in $P(B|A)^{n-1}$.

Such interweaving, we believe, would select and predict the nominal/previous computation value in order to qualify the difference of it.

Nevertheless, we find that the 'imaginary probability' of $P(A|Bi)$ reduces computation to an independent cognition of $i = 1$, or to $i^2 = -1$. And whether any supposition in any $P(A|Bi)$ is 'FALSE' or 'TRUE', it is for Bayesian inference computation to consider for.

5. Results

1. We prove that the time occurrence and variance of $P(A|B)$ at the moment of contemplation could be more narrowed via $iPt = \frac{A}{tv} = \frac{iPt(B)}{A}$, which signifies certain relativity of Bayesian inference even by standards of automated equations.

2. The bound limitation of [X, Y, Z] variance should construct the time (t) value into a time preference of the observer, in where we set $\int \int \int \frac{AvXn}{t \in A(Z,Y)}$, which is computable. Therefore, the bound set of [X, Y, Z] in such inference is undeniable.

3. Imaginary unit of i^2 may be transferable into i(B)(A), and may be computable as i in transfer of $i^2 \rightarrow i$, which leads us to believe that negative number is negative only abstractly, nevertheless, could be reversed and applied into positive value $i \rightarrow \mathbb{R}$.

4. The data proof and data analysis is comparative and interchangeable, as soon as it goes by non-accordance and exclusion in $iP \rightarrow P = \sqrt{-1} = \frac{1}{x} = p(A)$ when its algebraically possible to shift from negative into positive.

[1] See more examples in: No Author, The GNU C Library, 20.9 Complex Numbers

5. The computation of $P(A|B) \rightarrow \frac{P(A|X)\left(B|Y^{-1}\right)P(A|Z)}{P(B|Y)Log-1}$ in computational programming may be possible in logical terms only.

6. The main computation of stochastic modeling applies to the free bound observation in:

$$iP(A|X) = \frac{P(B)_i(P(A)}{P(B)}$$

in where we yield similar to Markov decimal data but in parallel of 2 or more possibilities in average index per item, A or A_i, and we get the both results in positive and negative values ($\mathbb{R}$).

7. The average index in:

$$indexA/A_i = \frac{P(A)(iP(A|X)}{iP}$$

and

$$indexB/B_i = \frac{P(B)(iP(B|Y)}{iP}$$

apply as contra-arguments of the Bayesian probability and permit i number probability to coexist with the factual one in $\mathbb{R}$. We presume that in some instances the iP even may take over the factual P, if it's in the average index $\geq$ over it.

5.1 Tables and Figures

An example of iP and P in differentiation.

Condition	Result index	Computation set	Computation ratio
$iP(A\|B)$	(0.4)	X_{n-1}	$\frac{X_{n-1}}{iy}$
$P(A\|B)$	(0.5)	X_n	$\frac{X_n}{P(B)}$

6. Discussion

If we reflect upon the current state of Bayesian inference which coexists in parallel with the statistical conclusion, we would stumble upon an artificial psychology of an actual and subjective choice selection.

Thence, we would never understand the difference between the human imagination and just 'imagining' the probability, unless we define imaginary integer as a 'possibility' index. Which is by far would yield less rational, but nevertheless, existent and accountable probability.

While having an artificial doubt or any other non-factual, hypothetical model, why would a machine collect such 'trifle' alongside with the factual data? Why would we complicate, if we may produce the result regardless of it and make it linear?

And it's the same question of independent cognition. The fact of 'A' shall be compared to the possibility of 'B', the solidity of '1' shall be juxtaposed to '-1'. A solid statement of A shall be shattered onto any forms of i ('B', '-1', etc) in order to re-solidify 'A' into a precedent, into a pattern of self-learning by 'imaginary mistakes'.

The deviation of a culprit is the teacher, and if we sift his doubts and strays through the solid logic of comparison, then we would reap the real diamonds.

Acknowledgments

Special gratitude to Terrence Nuss and to the Tatarstan Republic.

References

Alessandro, C., Massimo, V., & Alberto, G. (2014). *Dynamic Programming and Bayesian Inference, Concepts and Applications*. Intech. http://dx.doi.org/10.5772/57005

Applebaum, D. (1996). *Probability and information: An integrated approach*. Cambridge University Press.

Anderson, D. F. (1991). *Continuous-time Markov chains*. Citeseer. http://dx.doi.org/10.1007/978-1-4612-3038-0

Claret, G., Rajamani, S. K., Nori, A. V., Gordon, A. D., & Borgström, J. (2013). *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering*. pp.92–102. ACM.

Corfield, D., & Williamson, J. (2001). *Foundations of Bayesianism.* vol. 24. Springer Science & Business Media. http://dx.doi.org/10.1007/978-94-017-1586-7

Diaconis, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthese, 36*(2), 1–2. Springer.

Nau, R. F. (2001). De Finetti was right: probability does not exist. *Theory and Decision, 51*(2-4), 89–124. Springer. http://dx.doi.org/10.1023/A:1015525808214

Noack, B., Klumpp, V., Brunn, D., & Hanebeck, U. D. (2008). *Nonlinear Bayesian estimation with convex sets of probability densities*. Information Fusion, 2008 11th International Conference. pp.1–8. IEEE.

Paisley, J., & Blei, D., & Jordan, M. (2012). *Variational Bayesian inference with stochastic search*. arXiv preprint arXiv:1206.6430

International Journal of Statistics and Probability; Vol. 4, No. 4; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Estimation Techniques for Regression Model with Zero-inflated Poisson Data

Shakhawat Hossain[1] & Hatem A. Howlader[1]

[1] Department of Mathematics and Statistics, University of Winnipeg, Winnipeg, Canada

Correspondence: Shakhawat Hossain, Associate Professor, Department of Mathematics and Statistics, University of Winnipeg, 515 Portage Avenue, Winnipeg, MB, Canada R3B 2E9. Phone: 1-204-786-9492. E-mail: sh.hossain@uwinnipeg.ca

Received: August 10, 2015 Accepted: August 24, 2015 Online Published: September 23, 2015

doi:10.5539/ijsp.v4n4p64 URL: http://dx.doi.org/10.5539/ijsp.v4n4p64

Abstract

Researchers in many fields including biomedical often make statistical inferences involving the analysis of count data that exhibit a substantially large proportion of zeros. Subjects in such research are broadly categorized into low-risk group that produces only zero counts and high-risk group leading to counts that can be modeled by a standard Poisson regression model. The aim of this study is to estimate the model parameters in presence of covariates, some of which may not have significant effects on the magnitude of the counts in presence of a large proportion of zeros. The estimation procedures we propose for the study are the pretest, shrinkage, and penalty when some of the covariates may be subject to certain restrictions. Properties of the pretest and shrinkage estimators are discussed in terms of the asymptotic distributional biases and risks. We show that if the dimension of parameters exceeds two, the risk of the shrinkage estimator is strictly less than that of the maximum likelihood estimator, and the risk of the pretest estimator depends on the validity of the restrictions on parameters. A Monte Carlo simulation study shows that the mean squared errors (MSE) of shrinkage estimator are comparable to the MSE of the penalty estimators and in particular it performs better than the penalty estimators when the dimension of the restricted parameter space is large. For illustrative purposes, the methods are applied to a real life data set

Keywords: adaptive LASSO, asymptotic bias and risk, LASSO, likelihood ratio test, shrinkage and pretest estimators, zero-inflated data.

1. Introduction

Data with large number of zeros are often encountered in many studies including medical and public health. Failure to account for the extra zeros may result in biased parameter estimation and misleading inference. In recent years the regression method for modelling such data has become very popular. It is based on the assumption that the response is generated by a mixture of a degenerate distribution at zero and a standard Poisson distribution. For example, the National Medical Expenditure Survey data contain an excessive number of zeros in the hospital admission, and the probability mass at point zero exceeds that allowed under a standard parametric family of distributions. We will analyze this survey data in Section 5 where the number of hospital admission ranges from 0 to 8, and 80.4% of the respondents have no hospital admission, which presents possible zero-inflation. The zero-inflated Poisson regression (ZIPR) model can be used to analyze this data which assumes that the population of respondents can be divided into two subpopulations. The first of which generates counts (number of hospital admissions) that follow a Poisson distribution with the usual parameter, λ and the second subpopulation generates only the zeros (no hospital admission) with probability, p that follows binomial distribution with logit model.

The ZIPR models were first introduced by Mullahy (1986) in the econometric literature. Their use has become very broad since the publication of the pioneering work by Lambert (1992) in which manufacturing defects were considered. Ridout et al. (1998) cited examples of data with too many zeros from various disciplines including medicine, agriculture, and the use of recreational facilities. Hall (2000) considered ZIPR and zero-inflated binomial models with random intercept, and provided an EM algorithm for model estimation. Hinde & Demetrio (1998) reviewed the literature of ZIPR model and cited examples from agriculture, econometrics, manufacturing, road safety, species abundance, and medical sciences. Jansakul & Hinde (2002) extended the van den Broek (1995) score test to the more general situation where the zero probability is allowed to depend on covariates. Feng & Zhu (2011) studied a semiparametric ZIP mixture model, and a Monte Carlo EM algorithm was provided for model estimation. For other aspects of ZIPR models, see among others, Dietz & Bohning (2000) and Ridout et al. (2001).

The motivation of the present paper comes from the need to be able to identify the significant covariates of the ZIPR model in presence of a higher proportion of zeros in the response variable. With this motivation, we consider an unrestricted model that includes all covariates and possible extraneous variables and a restricted model that includes some covariates of concern while leaving out extraneous variables. One way to deal with this framework is to use the pretest procedures that test whether the coefficients of the extraneous variables are zero and then estimate the parameters in the model that include coefficients that are rejected by the test. Another approach is to use Stein type shrinkage estimators where the estimated regression coefficient vector is shrunk in the direction of linear restriction among the parameters. Sapra (2003) considered the pretest method to estimate the parameters of the Poisson regression model. For the properties of pretest and shrinkage estimation strategies for parametric and semiparametric linear models, see Nkurunziza (2013), Hossain et al. (2012), and Liang & Song (2009), among others.

This paper also studies the penalty estimators such as, least absolute shrinkage and selection operator (LASSO) and adaptive LASSO and compares their performances in terms of biases and mean squared errors with the shrinkage and pretest estimators through a simulation study. The LASSO was developed by Tibshirani (1996) for simultaneous variable selection and parameter estimation. Unlike LASSO, the adaptive LASSO proposed in Zou (2006) permits different weights for different parameters. The adaptive LASSO has been shown to have the oracle property, that is, consistency in variable selection and asymptotic normality. Hossain & Ahmed (2012) considered the shrinkage and penalty methods for estimating the Poisson regression model when some of the covariates may be inactive for the response. Zeng et al. (2014) proposed a variable selection approach for ZIPR models via adaptive LASSO.

The rest of this paper is organized as follows. The model and suggested estimators are introduced in Section 2. The asymptotic properties of the proposed estimators and their asymptotic distributional biases and risks are presented in Section 3. The results of a simulation study are reported in Section 4. Application to real life data and a comparison of our methods are described in Section 5. Finally, concluding remarks are given in Section 6.

2. Models and the Proposed Estimators

2.1 Models and Estimation Method

Suppose that the counts, Y_i, $i = 1, 2, \cdots, n$ are generated independently according to a zero-inflated Poisson distribution. The zeros are assumed to arise in two distinct underlying processes. The first process occurs with probability p_i and produces only zeros, while the second process occurs with probability $1 - p_i$ and leads to a standard Poisson model with parameter λ_i and hence a chance of further zeros. In general, the inevitable zeros from the first process are called structural zeros and those from the Poisson process are called sampling zeros, see for example, Ridout et al. (1998) and Jansakul & Hinde (2002). These two processes give rise to a simple two-component mixture distribution with probability mass function

$$Pr(Y = y_i) = \begin{cases} p_i + (1 - p_i)exp(-\lambda_i), & y_i = 0 \\ (1 - p_i)\frac{exp(-\lambda_i)\lambda^{y_i}}{y_i!}, & y_i \geq 1, \quad 0 \leq p_i \leq 1 \end{cases} \tag{1}$$

and $E(Y_i) = (1 - p_i)\lambda_i$, $var(Y_i) = \lambda_i + \left(\frac{p_i}{1-p_i}\right)\lambda_i^2$. The ZIPR model reduces to a Poisson regression model when $p_i = 0$, and exhibits overdispersion when $p_i > 0$. Lambert (1992) incorporated covariates using a log link for λ_i and logit link for p_i.

$$log(\lambda_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} \quad \text{and} \quad \log\left(\frac{p_i}{1 - p_i}\right) = \boldsymbol{z}_i^\top \boldsymbol{\gamma}, \tag{2}$$

where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{ip})^\top$ and $\boldsymbol{z}_i = (z_{i1}, z_{i2}, \cdots, z_{iq})^\top$ are the covariate vectors, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the $p \times 1$ and $q \times 1$ vectors of unknown regression parameters. Let us assume $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$. The covariates that affect the Poisson parameter of the first process may or may not be the same that affect the probability of the second process. When the covariates are the same and λ_i and p_i are not functionally related, then $\boldsymbol{x}_i = \boldsymbol{z}_i$ and in that case the ZIPR involves twice as many parameters as the Poisson regression. In model (2), we are interested in testing the following hypothesis:

$$H_0 : \boldsymbol{H\theta} = \boldsymbol{h} \quad \text{versus} \quad H_A : \boldsymbol{H\theta} \neq \boldsymbol{h},$$

where $\boldsymbol{H}$ is an $r \times (p + q)$ matrix of full rank and $\boldsymbol{h}$ is a $r \times 1$ vector of constant terms.

For a random sample, $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)$, the log-likelihood function is given by

$$\begin{aligned}
\ell = \ell(\boldsymbol{\lambda}, \boldsymbol{p}; \boldsymbol{y}) &= \sum_{i=1}^{n} \Big\{ I_{(y_i=0)} \ln[p_i + (1-p_i)\exp(-\lambda_i)] + I_{(y_i>0)}[\ln(1-p_i) - \lambda_i + y_i \ln\lambda_i - \ln(y_i!)] \Big\} \\
&= \sum_{i=1}^{n} \Big\{ I_{(y_i=0)} \ln[\exp(z_i^\top \boldsymbol{\gamma}) + \exp(-\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}))] \\
&+ I_{(y_i>0)}[y_i \boldsymbol{x}_i^\top \boldsymbol{\beta} - \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \ln(y_i!)] \Big\} - \sum_{i=1}^{n} \ln(1 + \exp(z_i^\top \boldsymbol{\gamma})). \qquad (3)
\end{aligned}$$

where $I_{(\cdot)}$ is an indicator function, which is equal to 1 if the event is true and 0 otherwise.

The log-likelihood function (3) of the ZIPR model is quite complicated, especially as the first term involves both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Also, the responses are from a mixture distribution that includes both sets of the parameters p_i and λ_i. The computation thus becomes quite challenging in terms of the variance-covariance matrix and accuracy when using the Newton-Raphson algorithm. And it is due to the additional number of parameters to be estimated for the proposed model from the complicated nature of the likelihood function (3). To avoid this complication, we use the EM algorithm to maximize the log-likelihood function, see, Hall (2000) and Lambert (1992).

The EM algorithm is based on a latent variable U_i, where we observe U_i as 1, when Y_i is from the perfect, zero state (or first process) and U_i as 0, when Y_i is from the Poisson state (or second process). To formulate the log-likelihood for the complete data, we use the conditional probability:

$$\begin{aligned}
& Pr(Y_i = y_i, U_i = u_i | \boldsymbol{x}_i, z_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) \\
= \; & Pr(Y_i = y_i | U_i = u_i, \boldsymbol{x}_i, z_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) \times Pr(U_i = u_i | \boldsymbol{x}_i, z_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) \\
= \; & \left(\frac{\exp(z_i^\top \boldsymbol{\gamma})}{1 + \exp(z_i^\top \boldsymbol{\gamma})} \right)^{u_i} \left(\frac{\exp(y_i \boldsymbol{x}_i^\top \boldsymbol{\beta} - \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}))}{y_i!(1 + \exp(z_i^\top \boldsymbol{\gamma}))} \right)^{1-u_i}
\end{aligned}$$

Thus, the complete log-likelihood based on $(\boldsymbol{Y}, \boldsymbol{U})$ is

$$\begin{aligned}
L_c(\boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{U}, \boldsymbol{x}_i, z_i) &= \ln \left[\prod_{i=1}^{n} Pr(Y_i = y_i, U_i = u_i | \boldsymbol{x}_i, z_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) \right] \\
&= \sum_{i=1}^{n} \{ u_i z_i^\top \boldsymbol{\gamma} - u_i \ln(1 + \exp(z_i^\top \boldsymbol{\gamma})) \} \\
&+ \sum_{i=1}^{n} (1-u_i) \left[(y_i \boldsymbol{x}_i^\top \boldsymbol{\beta} - \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \ln(y_i!) - \ln(1 + \exp(z_i^\top \boldsymbol{\gamma}))) \right] \\
&= L_{c1}(\boldsymbol{\gamma}; \boldsymbol{U}, \boldsymbol{x}_i, z_i) + L_{c2}(\boldsymbol{\beta}; \boldsymbol{U}, \boldsymbol{x}_i, z_i) - \sum_{i=1}^{n} (1-u_i) \ln(y_i!)), \qquad (4)
\end{aligned}$$

where $L_{c1} = \sum_{i=1}^{n} \{ u_i z_i^\top \boldsymbol{\gamma} - \ln(1 + \exp(z_i^\top \boldsymbol{\gamma})) \}$, $L_{c2} = \sum_{i=1}^{n} (1-u_i)(y_i \boldsymbol{x}_i^\top \boldsymbol{\beta} - \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}))$, and $U = \{ u_i; i = 1, 2, \cdots, n \}$.

To implement the EM algorithm, we first initialize $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ by letting $U_i = I_{(y_i=0)}$. In the E-step, we use the initial values of $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ to calculate the expectation of U_i, and use it as the estimator of U_i. In the M-step, we use the estimate U_i to maximize $L_c(\boldsymbol{\beta}, \boldsymbol{\gamma}; \boldsymbol{U}, \boldsymbol{x}_i, z_i)$, which gives the unrestricted maximum likelihood estimator for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. The iteration $(k+1)$ of the EM algorithm requires the following steps.

E-Step: Estimate $U_i^{(l)}$ by using the means given $\boldsymbol{\gamma}^{(l)}$ and $\boldsymbol{\beta}^{(l)}$,

$$\begin{aligned}
U_i^{(l)} &= E(U_i | y_i, \gamma^{(l)}, \boldsymbol{\beta}^{(l)}) \\
&= E(U_i = 1 | y_i, \gamma^{(l)}, \boldsymbol{\beta}^{(l)}) \\
&= \frac{Pr(Y_i = y_i | U_i = 1) Pr(U_i = 1)}{Pr(Y_i = y_i | U_i = 1) Pr(U_i = 1) + Pr(Y_i = y_i | U_i = 0) Pr(U_i = 0)} \\
&= \left[1 + \exp(-\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}^{(l)}) - z_i^\top \boldsymbol{\gamma}^{(l)}) \right]^{-1}, \text{ if } y_i = 0, \text{ and } 0, \text{ if } y_i \geq 1.
\end{aligned}$$

M-Step: Given $U_i = U_i^{(l)}$, maximize $L_{c1}(\boldsymbol{\gamma}; \boldsymbol{U}^{(l)}, \boldsymbol{x}_i, \boldsymbol{z}_i)$ and $L_{c2}(\boldsymbol{\beta}; \boldsymbol{U}^{(l)}, \boldsymbol{x}_i, \boldsymbol{z}_i)$ with respect to $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$, respectively:

$$\boldsymbol{\gamma}^{(l+1)} = \underset{\boldsymbol{\gamma}}{\operatorname{argmin}} \left\{ -L_{c1}(\boldsymbol{\gamma}; \boldsymbol{U}^{(l)}, y_i, \boldsymbol{x}_i, \boldsymbol{z}_i) \right\} \tag{5}$$

$$\boldsymbol{\beta}^{(l+1)} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ -L_{c2}(\boldsymbol{\beta}; \boldsymbol{U}^{(l)}, y_i, \boldsymbol{x}_i, \boldsymbol{z}_i) \right\} \tag{6}$$

The iteration stops when (5) and (6) converge simutaneously. The final estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\gamma}}^\top)^\top$ is known as the unrestricted maximum likelihood estimator (UMLE).

Let $I_{(\boldsymbol{\beta},\boldsymbol{\gamma})}$ be the information matrix of the estimator $\hat{\boldsymbol{\theta}}$. If $\frac{1}{n} I_{(\boldsymbol{\beta},\boldsymbol{\gamma})}$ has a positive definite limit satisfying some regularity conditions, as in the work of McCullagh (1983), the quantity $\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)$ is asymptotically normally distributed with mean vector $\mathbf{0}$ and information matrix $I^{-1}_{(\boldsymbol{\beta},\boldsymbol{\gamma})}$ (Lambert, 1992). The matrix $I_{(\boldsymbol{\beta},\boldsymbol{\gamma})}$ can be partitioned as

$$\begin{pmatrix} I_{\boldsymbol{\beta},\boldsymbol{\beta}} & I_{\boldsymbol{\beta},\boldsymbol{\gamma}} \\ I_{\boldsymbol{\gamma},\boldsymbol{\beta}} & I_{\boldsymbol{\gamma},\boldsymbol{\gamma}} \end{pmatrix},$$

where the elements $I_{\boldsymbol{\beta},\boldsymbol{\beta}}$, $I_{\boldsymbol{\beta},\boldsymbol{\gamma}} = I^\top_{\boldsymbol{\gamma},\boldsymbol{\beta}}$, and $I_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$ are, respectively,

$$-\mathrm{E}\left[\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\right], -\mathrm{E}\left[\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}}\right], \text{ and } -\mathrm{E}\left[\frac{\partial^2 l}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top}\right]$$

with

$$\begin{aligned}
\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} &= \sum_{i=1}^{n} \left\{ I_{(y_i=0)} \left[\frac{-\exp(-\lambda_i)[(1-\lambda_i)p_i + (1-p_i)\exp(-\lambda_i)](1-p_i)\lambda_i}{[p_i + (1-p_i)\exp(-\lambda_i)]^2} \right] \right. \\
&+ \left. I_{(y_i>0)}(-\lambda_i) \right\} x_{ij} x_{jk}, \\
\frac{\partial^2 l}{\partial \beta_j \partial \gamma_r} &= \frac{\partial^2 l}{\partial \gamma_j \partial \beta_r} = \sum_{i=1}^{n} \left\{ I_{(y_i=0)} \left[\frac{\lambda_i exp(-\lambda_i)}{[p_i + (1-p_i)\exp(-\lambda_i)]^2} \right] \right\} x_{ij} z_{ir}, \\
\frac{\partial^2 l}{\partial \gamma_r \partial \gamma_s} &= \sum_{i=1}^{n} \left\{ I_{(y_i=0)} \left[\frac{-(1-\exp(-\lambda_i))^2}{[p_i + (1-p_i)\exp(-\lambda_i)]^2} \right] + I_{(y_i>0)} \left[\frac{-1}{(1-p_i)^2} \right] \right\} z_{ir} z_{is},
\end{aligned}$$

where $j, k = 1, 2, \cdots, p$ and $r, s = 1, 2, \cdots, q$.

Suppose now that our interest centers in estimating the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ from (4) under the linear restriction $\boldsymbol{H\theta} = \boldsymbol{h}$. The steps of the EM-algorithm for estimating the parameters using log-likelihood (4) under the above restriction are the same. We simply replace $\boldsymbol{\theta}^{(l+1)} = (\boldsymbol{\beta}^{(l+1)}, \boldsymbol{\gamma}^{(l+1)})$ by $\boldsymbol{\theta}_c^{(l+1)} = (\boldsymbol{\beta}_c^{(l+1)}, \boldsymbol{\gamma}_c^{(l+1)})$ in the M-step. The resulting estimator, $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}^\top, \tilde{\boldsymbol{\gamma}}^\top)^\top$ is known the restricted maximum likelihood estimator (RMLE).

The likelihood ratio test statistic can be used to test $H_0 : \boldsymbol{H\theta} = \boldsymbol{h}$ vs. $H_a : \boldsymbol{H\theta} \neq \boldsymbol{h}$. If $\tilde{\boldsymbol{\theta}}$ maximizes the log likelihood of the ZIPR model under H_0 of dimension $k - r$, where $k = p + q$ and $\hat{\boldsymbol{\theta}}$ maximizes the log likelihood of the ZIPR model under a nested alternative hypothesis H_A of dimension k, then the test statistic D_n is

$$\begin{aligned}
D_n &= 2[l(\hat{\boldsymbol{\theta}}) - l(\tilde{\boldsymbol{\theta}}] \\
&= \left(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h}\right)^\top [\boldsymbol{H}\mathrm{var}(\hat{\boldsymbol{\theta}})\boldsymbol{H}^\top]^{-1} \left(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h}\right) + o_p(1)
\end{aligned} \tag{7}$$

Under H_0 and suitable regularity conditions, D_n is asymptotically distributed as chi-square with r degrees of freedom (Lambert, 1992). So that a likelihood ratio test can be performed using approximate critical values from the chi-square distribution.

2.2 The Pretest and Shrinkage Estimators

The pretest and shrinkage estimators are based on the test statistic D_n of (7) for testing $H_0 : \boldsymbol{H\theta} = \boldsymbol{h}$. Specifically, the pretest estimator (PT) of $\boldsymbol{\theta}$ is defined as

$$\hat{\boldsymbol{\theta}}_{PT} = \hat{\boldsymbol{\theta}} - (\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}) I(D_n \leq \chi^2_{r,\alpha}),$$

where $I(A)$ is an indicator function of a set A, and $\chi^2_{r,\alpha}$ is the α-level critical value of the approximate distribution of D_n under H_0. From the above definition, one can see that if the data yield $D_n < \chi^2_{r,\alpha}$, then $\hat{\boldsymbol{\theta}}_{PT} = \tilde{\boldsymbol{\theta}}$, otherwise

$\hat{\boldsymbol{\theta}}_{PT} = \hat{\boldsymbol{\theta}}$. So the PT is indeed a simple mixture of the UMLE and RMLE estimators. In an ordinary two-step procedure, one would test the hypothesis $H_0 : \boldsymbol{H\theta} = \boldsymbol{h}$ first, then based on the test result decide which estimator should be adopted. The PT simply combines these two steps to form a single one. That is, testing and estimation are done simultaneously. For details, see Hossain et al. (2009) and Ahmed at al. (2006), and others.

Because of extreme choices for either the UMLE or RMLE, the pretest procedures are not admissible for some models, even though they may improve upon the UMLE, a well-documented fact in the literature (Judge & Bock, 1978). This motivates us to define a shrinkage estimator, which is a smoothed version of $\hat{\boldsymbol{\theta}}_{PT}$:

$$\hat{\boldsymbol{\theta}}_{SE} = \tilde{\boldsymbol{\theta}} + \left(1 - (r-2)D_n^{-1}\right)(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}), \quad r \geq 3. \tag{8}$$

This estimator is a weighted average of UMLE $\hat{\boldsymbol{\theta}}$ and RMLE $\tilde{\boldsymbol{\theta}}$, where the weights are a function of the test statistic for testing $H_0 : \boldsymbol{H\theta} = \boldsymbol{h}$.

We note that when the test statistic D_n is very small in comparison with $r-2$, i.e., when the ratio $(r-2)/D_n$ is greater than one in absolute value, the shrinkage estimator $\hat{\boldsymbol{\theta}}_{SE}$ tends to shrink $\hat{\boldsymbol{\theta}}$ overly towards $\tilde{\boldsymbol{\theta}}$ and reversing the sign of $\hat{\boldsymbol{\theta}}$. Replacing $\left(1 - (r-2)D_n^{-1}\right)$ by $\left(1 - (r-2)D_n^{-1}\right)_+$ in (8), where $(x)_+ = x1_{(x\geq 0)}$, the positive-part shrinkage estimator, $\hat{\boldsymbol{\theta}}_{PSE}$ rectifies this problem. For details, see, Ahmed & Fallahpour (2012).

2.3 Penalty Estimators: LASSO and Adaptive LASSO

Let $\boldsymbol{\theta}$ be the parameter of interest and $\ell(\boldsymbol{\theta})$ be the log-likelihood function of the ZIPR, then the LASSO estimator (Tibshirani, 1996) is given by

$$\hat{\boldsymbol{\theta}}_{lasso}(\tau) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\{ -\ell(\boldsymbol{\theta}) + \tau \sum_{g=1}^{p+q} |\boldsymbol{\theta}_g| \right\},$$

where $\tau \geq 0$ is a penalization parameter controlling the amount of shrinkage on the regression coefficients. However, the LASSO applies same penalty to all the coefficients, which over-penalizes the important ones and accordingly results in biased estimators (Zou, 2006).

The adaptive LASSO (ALASSO) (Zou, 2006) offers an effective way to minimize this bias. It has been shown that the ALASSO enjoys the oracle property that is, the ALASSO is consistent in variable selection, and its estimators are asymptotically normal and unbiased. More explicitly, it assigns a higher weight to the small coefficients and lower weight to the large coefficients. The adaptive lasso estimator is defined as

$$\hat{\boldsymbol{\theta}}_{alasso}(\tau) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\{ -\ell(\boldsymbol{\theta}) + \tau \sum_{g=1}^{p+q} \boldsymbol{w}_g |\boldsymbol{\theta}_g| \right\}, \tag{9}$$

where $\boldsymbol{w} = (w_1, w_2, \cdots, w_{p+q})$ are adaptive weights, which are usually set to $\hat{\boldsymbol{w}} = 1/|\hat{\boldsymbol{\theta}}|$, here $\hat{\boldsymbol{\theta}}$ are the maximum likelihood estimators. The ALASSO estimator can be obtained by minimizing penalized log-likelihood function. First, we expand the likelihood function based on a Taylor series expansion at $\hat{\boldsymbol{\theta}}$:

$$\ell(\boldsymbol{\theta}) \approx \ell(\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top \ell'(\hat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top \ell''(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \tag{10}$$

where ℓ' and ℓ'' are the first and second derivatives, respectively. Since $\ell(\hat{\boldsymbol{\theta}})$ is constant and $\ell'(\hat{\boldsymbol{\theta}}) = 0$, so the penalized log-likelihood function is equivalent to

$$\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top \left[-\ell''(\hat{\boldsymbol{\theta}})\right](\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \sum_{g=1}^{p+q} |\boldsymbol{\theta}_g|. \tag{11}$$

Let $\hat{\boldsymbol{\Sigma}}$ be the estimated variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, which can be obtained from one of many statistical packages. And it provides an estimate of the Hessian matrix ℓ'', given as $\ell'' = -I_{\hat{\theta}} = -\hat{\boldsymbol{\Sigma}}^{-1}$. Here we use the *R* statistical package *pscl* (Jackman, 2012) in order to calculate $\hat{\boldsymbol{\Sigma}}$. Then we construct a pseudo-data set as

$$\boldsymbol{X}^* = \hat{\boldsymbol{\Sigma}}^{-1/2}, \boldsymbol{Y}^* = \hat{\boldsymbol{\Sigma}}^{-1/2}\hat{\boldsymbol{\theta}}.$$

The order of the square matrix $\boldsymbol{X}^*$ is $p+q+2$ and $\boldsymbol{Y}^*$ is a vector corresponding to $\boldsymbol{X}^*$. The ALASSO estimator (9) can now be re-written as

$$\hat{\boldsymbol{\theta}}_{alasso}(\tau) \approx \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\{ \frac{1}{2}(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\theta})^\top(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\theta}) + \tau \sum_{g=1}^{p+q} \boldsymbol{w}_g|\boldsymbol{\theta}_g| \right\}, \tag{12}$$

The estimator in equation (12) is similar to the least squares estimator with adaptive LASSO penalization (Zou, 2006). Various efficient algorithms, such as the least angle regression algorithm (Efron et al., 2004), the predictor-corrector algorithm (Park & Haste, 2007), and the coordinate descent algorithm (Friedman et al., 2010) can be used to conduct the minimization of Equation (12). Here we adopt the predictor-corrector algorithm which computes solutions along the entire penalization path of the coefficient estimates as τ varies. Starting at $\tau = \tau_{max}$, where τ_{max} is the largest τ that makes all the coefficients of $\hat{\boldsymbol{\theta}}(\tau)$ nonzero. This algorithm computes a series of solutions, each time estimating the coefficients with a smaller τ. The penalization parameter τ is selected using k-fold cross validation. For each fold, we obtain a series of models based on Schwarz's Bayesian criterion, commonly referred to as Bayesian information criterion corresponding to the values of τ and compute the right side of (12) using the omitted fold. Then we choose the value of τ for which the average cross-validation in the right side of (12) is minimized.

In the case of pretest and shrinkage estimators, it is possible to determine theoretically when they have smaller asymptotic risks than the unrestricted MLE. Similar results are not available for LASSO and adaptive LASSO estimators.

3. Asymptotic Results

In this section we study the asymptotic behavior of the proposed estimators and develop the results in terms of $\hat{\boldsymbol{\theta}}^*$, which could be any one of the estimators considered in this paper: $\hat{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}_{PT}$, $\hat{\boldsymbol{\theta}}_{SE}$, and $\hat{\boldsymbol{\theta}}_{PSE}$. In order to proceed with the asymptotic results, we first partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, where $\boldsymbol{\theta}_2$ is the coefficient vector of the inactive covariates for the ZIPR model. The main focus here is to evaluate the performance of these estimators when $\boldsymbol{\theta}_2$ is close to the null vector. To derive any meaningful results we consider a sequence of local alternatives

$$K_n : \boldsymbol{H}\boldsymbol{\theta} = \boldsymbol{h} + \frac{\boldsymbol{\omega}}{\sqrt{n}}, \tag{13}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \cdots, \omega_r)^\top \in \Re^r$ is a given vector of real numbers. In this framework, $\boldsymbol{\theta}$= $(\boldsymbol{\theta}_1^\top, \boldsymbol{0}^\top)^\top$, and the quantity $\frac{\omega}{\sqrt{n}}$ is the magnitude of the distance between the true model and the restricted model. For any fixed $\boldsymbol{\omega}$, this distance shrinks as the sample size increases.

To study the asymptotic risks (ADR) of the estimators, we define a quadratic loss function by using a positive definite matrix $\boldsymbol{W}$, namely

$$\mathcal{L}(\hat{\boldsymbol{\theta}}^*; \boldsymbol{W}) = \left[\sqrt{n}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})\right]' \boldsymbol{W} \left[\sqrt{n}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})\right],$$

where $\hat{\boldsymbol{\theta}}^*$ is any one of the estimators. The usual quadratic loss is defined when $\boldsymbol{W}$ is chosen as $\boldsymbol{I}$, the identity matrix. A general $\boldsymbol{W}$ gives a loss function that weighs differently for different $\boldsymbol{\theta}$'s.

If $\boldsymbol{V}$ is the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}^*$, the ADR of $\sqrt{n}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})$ is given by tr(WV). We assume that the cumulative distribution function of $\hat{\boldsymbol{\theta}}^*$ under K_n exists and can be denoted as

$$F(\boldsymbol{x}) = \lim_{n\to\infty} P\left[\sqrt{n}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}) \le \boldsymbol{x} | K_n\right],$$

where $F(\boldsymbol{x})$ is a nondegenerate distribution function. The ADR of $\hat{\boldsymbol{\theta}}^*$ is then defined as

$$\begin{aligned} R(\hat{\boldsymbol{\theta}}^*; \boldsymbol{W}) &= \int \cdots \int \boldsymbol{x}'\boldsymbol{W}\boldsymbol{x}\, dF(\boldsymbol{x}) \\ &= \text{trace}(\boldsymbol{W}\boldsymbol{V}), \end{aligned} \tag{14}$$

where $\boldsymbol{V} = \int \cdots \int \boldsymbol{x}\boldsymbol{x}' dG(\boldsymbol{x})$ is the dispersion matrix for the distribution function $F(\boldsymbol{x})$.

The shrinkage estimators are, in general biased, the bias, however is accompanied by a reduction in variance. The asymptotic distributional bias (ADB) of an estimator $\hat{\boldsymbol{\theta}}^*$ is defined as

$$\text{ADB}(\boldsymbol{\theta}^*) = \lim_{n\to\infty} E\left\{n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})\right\}.$$

Under the local alternatives (13), the following theorems help the derivation and numerical computation of the ADB and the ADR of the estimators.

Theorem 3.1 *If $\boldsymbol{I}_{\beta,\gamma}$ is nonsingular and $\Delta = \boldsymbol{\omega}^\top(\boldsymbol{H}\boldsymbol{I}^{-1}_{\beta,\gamma}\boldsymbol{H}^\top)^{-1}\boldsymbol{\omega}$, then under the local alternatives K_n in (13) and regularity conditions, we have as $n \to \infty$*

1. $\sqrt{n}(\boldsymbol{H}\hat{\boldsymbol{\theta}} - \boldsymbol{h}) \xrightarrow{\mathcal{L}} N(\boldsymbol{\omega}, \boldsymbol{H}\boldsymbol{I}^{-1}_{\beta,\gamma}\boldsymbol{H}^\top)$.
2. *The test statistic D_n in (7) converges to a non-central chi-squared distribution $\chi^2_r(\Delta)$ with r degrees of freedom and non-centrality parameter Δ.*

For a concise statement of the results, we use the following notations:

$$
\begin{aligned}
&\boldsymbol{\zeta} = \boldsymbol{I}^{-1}\boldsymbol{H}^\top(\boldsymbol{H}\boldsymbol{I}^{-1}\boldsymbol{H}^\top)^{-1}\\
&g_{r+i}(\Delta) = H_{r+i}(r-2,\Delta), \quad h_{j,r+i}(\Delta) = E\left(\chi^{-2j}_{r+i}(\Delta)\right)\\
&k_{j,r+i}(\Delta) = H_{r+i}(r-2,\Delta) + (r-2)^j E\left(\chi^{-2j}_{r+i}(\Delta) I(\chi^2_{r+i})(\Delta) < r-2\right), \quad \text{and}\\
&O_{j,r+i}(\Delta) = E\left((1-(r-2)E(\chi^{-2j}_{r+i}(\Delta)))^2 I(\chi^2_{r+i})(\Delta) < r-2\right)
\end{aligned}
$$

where $i = 2, 4$, $j = 1, 2$ and for simplicity we assume that $\boldsymbol{I}_{\beta,\gamma} = \boldsymbol{I}$. Here $H_{r+2}(\cdot,\Delta)$ is the distribution function of the $\chi^2_r(\Delta)$ distribution.

Theorem 3.2 *Using the definition of ADB and Theorem 3.1, the ADBs of the estimators are,*

$$
\begin{aligned}
ADB(\hat{\boldsymbol{\theta}}) &= \boldsymbol{0}, \;\; ADB(\tilde{\boldsymbol{\theta}}) = -\boldsymbol{\zeta}\boldsymbol{\omega}, \;\; ADB(\hat{\boldsymbol{\theta}}_{PT}) = -\boldsymbol{\zeta}\boldsymbol{\omega} g_{r+2}(\Delta)\\
ADB(\hat{\boldsymbol{\theta}}_{SE}) &= -(r-2)\boldsymbol{\zeta}\boldsymbol{\omega} h_{1,r+2}(\Delta), ADB(\hat{\boldsymbol{\theta}}_{PSE}) = ADB(\hat{\boldsymbol{\theta}}_{SE}) - \boldsymbol{\zeta}\boldsymbol{\omega} k_{1,r+2}(\Delta).
\end{aligned}
$$

The outline of the proof is similar to that of Theorem 3.1 of Ahmed & Fallahpour (2012).

Clearly, the bias of the estimators is a function of Δ. Therefore, for bias comparisons, it suffices to compare the scalar factor Δ only. It is clear that the ADB of RMLE is an unbounded function of Δ. The ADB($\hat{\boldsymbol{\theta}}_{SE}$) and ADB($\hat{\boldsymbol{\theta}}_{PSE}$) start from the origin, and as Δ increases, they increase to a maximum and then decrease to 0. Note that, $h_{1,r+2}(\Delta)$ is a decreasing log-convex function of Δ and the ADB of $\hat{\boldsymbol{\theta}}_{PT}$ is a function of Δ and α. For a fixed α, ADB($\hat{\boldsymbol{\theta}}_{PT}$) starts at zero, increases to a point, then decreases gradually to zero.

Theorem 3.3 *Under the local alternatives K_n in (13) and the assumptions of Theorem 3.1, the ADRs of the estimator are*

$$
\begin{aligned}
&ADR(\hat{\boldsymbol{\theta}}; \boldsymbol{W}) = tr(\boldsymbol{W}\boldsymbol{I}^{-1})\\
&ADR(\tilde{\boldsymbol{\theta}}; \boldsymbol{W}) = ADR(\hat{\boldsymbol{\theta}}; \boldsymbol{W}) - tr(\boldsymbol{W}\boldsymbol{\zeta}\boldsymbol{H}\boldsymbol{I}^{-1}) + \boldsymbol{\omega}^\top(\boldsymbol{\zeta}^\top\boldsymbol{W}\boldsymbol{\zeta})\boldsymbol{\omega}\\
&ADR(\hat{\boldsymbol{\theta}}_{PT}; \boldsymbol{W}) = ADR(\hat{\boldsymbol{\theta}}; \boldsymbol{W}) - g_{r+2}(\Delta) tr(\boldsymbol{W}\boldsymbol{\zeta}\boldsymbol{H}\boldsymbol{I}^{-1}) + \boldsymbol{\omega}^\top(\boldsymbol{\zeta}^\top\boldsymbol{W}\boldsymbol{\zeta})\boldsymbol{\omega}(2g_{r+2}(\Delta) - g_{r+4}(\Delta))\\
&ADR(\hat{\boldsymbol{\theta}}_{SE}; \boldsymbol{W}) = ADR(\hat{\boldsymbol{\theta}}; \boldsymbol{W}) + ((r-2)^2 h_{2,r+2}(\Delta) - 2(r-2)h_{1,r+2}(\Delta)) tr(\boldsymbol{W}\boldsymbol{\zeta}\boldsymbol{H}\boldsymbol{I}^{-1})\\
&+ \quad \boldsymbol{\omega}^\top(\boldsymbol{\zeta}^\top\boldsymbol{W}\boldsymbol{\zeta})\boldsymbol{\omega}((r-2)^2 h_{2,r+4}(\Delta) + 2(r-2)h_{1,r+2}(\Delta) - 2(r-2)h_{1,r+2}(\Delta))\\
&ADR(\hat{\boldsymbol{\theta}}_{PSE}; \boldsymbol{W}) = ADR(\hat{\boldsymbol{\theta}}_{SE}; \boldsymbol{W}) - O_{1,r+2}(\Delta) tr(\boldsymbol{W}\boldsymbol{\zeta}\boldsymbol{H}\boldsymbol{I}^{-1}) - O_{1,r+4}(\Delta)\boldsymbol{\omega}^\top(\boldsymbol{\zeta}^\top\boldsymbol{W}\boldsymbol{\zeta})\boldsymbol{\omega}\\
&+ \quad \frac{1}{2}(3g_{r+2}(\Delta) + k_{1,r+2}(\Delta))\boldsymbol{\omega}^\top(\boldsymbol{\zeta}^\top\boldsymbol{W}\boldsymbol{\zeta})\boldsymbol{\omega}.
\end{aligned}
$$

The outline of the proof is similar to that of Theorem 3.2 of Ahmed & Fallahpour (2012).

By comparing the ADRs of the estimators, we see that, as Δ moves away from 0, the risk of $\tilde{\boldsymbol{\theta}}$ becomes unbounded. That is, the RMLE $\tilde{\boldsymbol{\theta}}$ dominates the unrestricted estimator at and near $\Delta = 0$. The risk of $\hat{\boldsymbol{\theta}}_{PSE}$ is asymptotically superior to $\hat{\boldsymbol{\theta}}_{SE}$ for all values of Δ, with strict inequality holds for some Δ. Thus, not only does $\hat{\boldsymbol{\theta}}_{PSE}$ confirm the inadmissibility of $\hat{\boldsymbol{\theta}}_{SE}$, but it also provides a simple superior estimator. Further, the largest risk improvement of $\hat{\boldsymbol{\theta}}_{PSE}$ over $\hat{\boldsymbol{\theta}}_{SE}$ is at and near the null hypothesis. Also, by comparing the ADRs of $\hat{\boldsymbol{\theta}}_{SE}$, $\hat{\boldsymbol{\theta}}_{PSE}$, and $\hat{\boldsymbol{\theta}}$, it can be easily shown that, under certain conditions ADR($\hat{\boldsymbol{\theta}}_{PSE}, \boldsymbol{W}$) $\leq$ ADR($\hat{\boldsymbol{\theta}}_{SE}, \boldsymbol{W}$) $\leq$ ADR($\hat{\boldsymbol{\theta}}; \boldsymbol{W}$) for all $\Delta \geq 0$. For a given α, PT is not uniformly better than the unrestricted estimator near the null hypothesis. One may determine an α such that PT has a minimum guaranteed risk. For any given $\boldsymbol{H}$, $\boldsymbol{I}$, and $\boldsymbol{h}$, the relative efficiency (inverse of risk) of PT

to UMLE is a function of α and λ. If the minimum efficiency required is RE_0, then we can choose α by solving the equation $\min_{\lambda}\{\text{Relative Efficiency}(\alpha, \Delta)\} = RE_0$. The exact solution may not be available, but one can use a numerical method to search for the minimum.

In order to explain and quantify the properties of the theoretical results, we conduct a simulation study to compare the performance of the suggested estimators.

4. Simulation Studies

The zero-inflated count responses are generated from the following mixture of two models:

$$log(\lambda_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta} \quad \text{and} \quad \log\left(\frac{p_i}{1-p_i}\right) = z_i^\top \boldsymbol{\gamma},$$

where $\boldsymbol{x}_i$ are generated from a multivariate normal distribution $N_p(0, \boldsymbol{\Sigma})$, with elements $\sigma_{ij} = \rho^{|i-j|}$ $(i, j = 1, 2, \cdots, 20)$ and z_i are generated from uniform, and Poisson distributions. We consider the correlation, $\rho = 0.50$ among the covariates, and sample sizes $n = 150$ and 350. For small samples, say $n = 50$, the likelihood function does not converge. The estimations of the variance-covariance matrix as well as the regression coefficients in such cases are unstable for the ZIPR model. We tried different sample sizes. In order for the likelihood function to converge, it needed a minimum sample of size 150. We set $\alpha = 0.05$ for the pretest method. Our objective is to test the hypothesis

$$H_0 : \boldsymbol{\theta}_2 = \boldsymbol{0} \quad \text{versus} \quad H_A : \boldsymbol{\theta}_2 \neq \boldsymbol{0}.$$

We set the true values of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top = \{(\boldsymbol{\beta}_1, \boldsymbol{\gamma}_1)^\top, (\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2)^\top\}$ as:

$$\boldsymbol{\theta} \quad = \quad ((1.1, -0.36, -0.36, 1.08, 0.92)^\top, -0.2^\top),\ \boldsymbol{b}^\top),$$

where $\boldsymbol{b} = \boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2, \boldsymbol{\gamma}_2)$ is a zero vector with different lengths. The number of replications is set to 1, 000 for all cases. We define the distance between the simulation model and the restricted model by

$$\Delta = \|\boldsymbol{\theta} - \boldsymbol{\theta}^{(0)}\|^2,$$

where $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^\top, \boldsymbol{0}^\top)^\top$, is a 6×1 true parameter vector in the simulation model and $\|\cdot\|$ is the Euclidian norm. Samples are generated using Δ between 0 and 1.

Based on the simulated data, we estimate the mean squared errors (MSE) of all the estimators studied in this paper. We calculate the simulated mean squared error (SMSE) by using the the empirical formula

$$SMSE(\boldsymbol{\theta}^*) = \sum_{i=1}^{p+q} (\theta_i^* - \theta_i)^2,$$

where $\hat{\boldsymbol{\theta}}^*$ as before denotes any one of the estimators $\hat{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}_{PT}$, $\hat{\boldsymbol{\theta}}_{SE}$, $\hat{\boldsymbol{\theta}}_{PSE}$, $\hat{\boldsymbol{\theta}}_{lasso}$, and $\hat{\boldsymbol{\theta}}_{alasso}$. The simulated relative efficiency (SRE) between two estimators is defined as $SRE(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}^*) = SMSE(\hat{\boldsymbol{\theta}})/SMSE(\hat{\boldsymbol{\theta}}^*)$, where $\hat{\boldsymbol{\theta}}$ is the unrestricted estimator, treated as benchmark. A value of $SRE(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}^*)$ greater than one indicates that $\hat{\boldsymbol{\theta}}^*$ performs better than $\hat{\boldsymbol{\theta}}$, and vice versa. Table 1 and Figure 1 give results for $\Delta \geq 0$ and (k_1, r)=(6, 3), (6, 6), (6, 11), (6, 14), where $k = p + q = (k_1, r)$. All analyses are performed using the statistical software *R*.

We observe from Table 1 and Figure 1 that when $\Delta = 0$, the RMLE outperforms all other estimators for both the cases considered. As Δ moves away from 0, the RMLE becomes unbounded. That is, the SRE curve for RMLE with respect to UMLE sharply decays and that it goes below the horizontal line when $\Delta > 0$. The positive-shrinkage estimator converges to 1 at a slower rate as Δ moves away from 0. This indicates that in the event of imprecise auxiliary information about the regression parameters, the PSE has the smallest risk among all other estimators for a range of Δ irrespective of the value of ρ. Pretest estimator dominates the shrinkage estimators when the restriction is nearly correct. Otherwise, it becomes unbounded at a faster rate than the RMLE. However, with the increase in Δ, the SRE of PT approaches 1 at some point from below. This suggests that neither PT nor RMLE is uniformly better when $\Delta > 0$.

The simulated relative efficiency of the shrinkage estimator typically lies above that of the UMLE, and is nearly as good as the RMLE at and near the null hypothesis $\Delta = 0$. This improvement relative to the RMLE is substantial for the increasing number of inactive covariates. Some readers may be surprised by the strong performance of

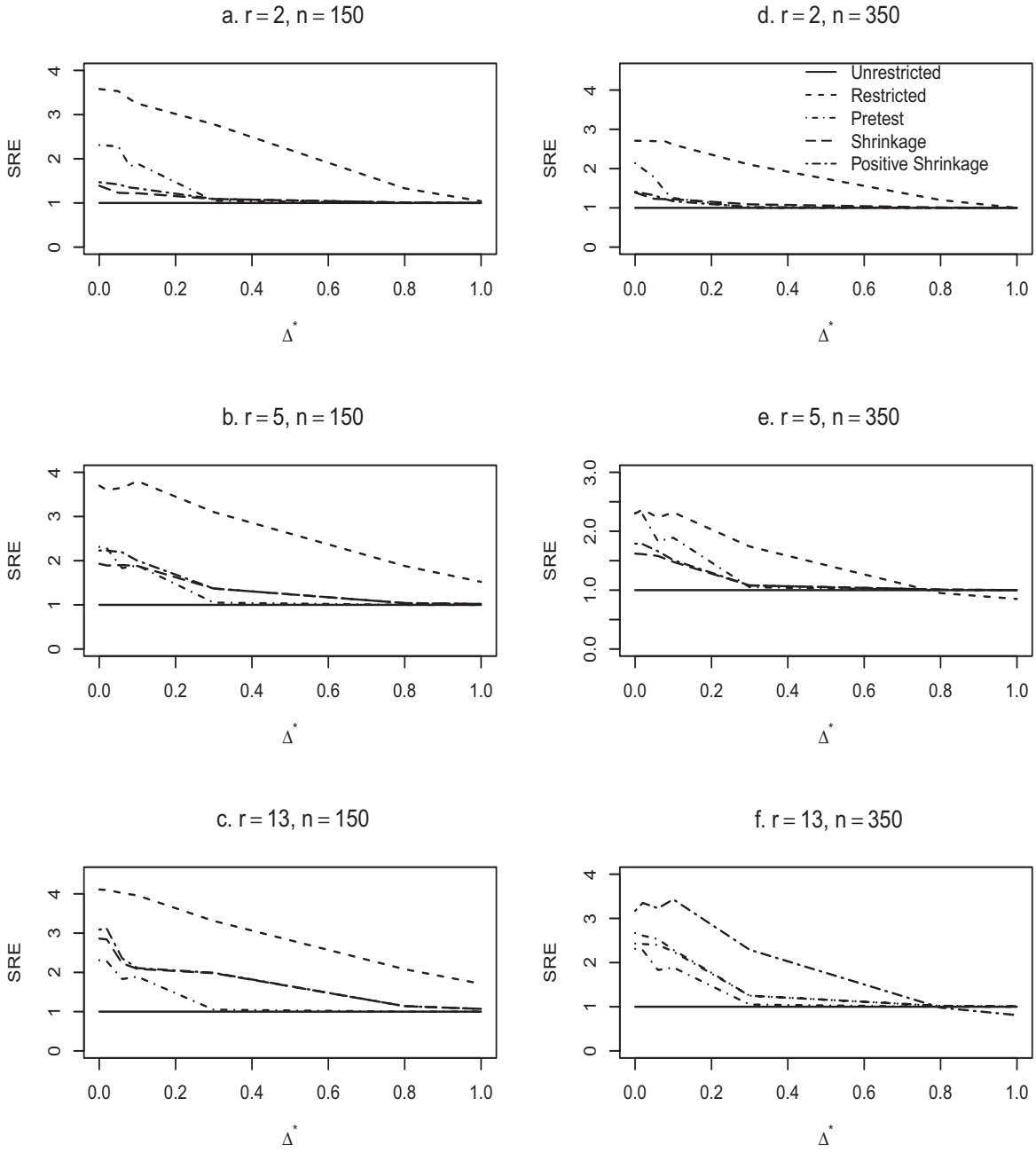


Figure 1. Simulated Relative efficiencies with respect to $\hat{\boldsymbol{\theta}}$ of the estimators for $\Delta \geq 0$. Here $k_1 = 6,\ r = 2, 10, 18$; $n = 150$ for the first column (Figures a-c); $k_1 = 6,\ r = 2, 10, 18$; $n = 350$ for the second column (Figures d-e).

the shrinkage estimators relative to the UMLE. However, this is precisely the lesson of Theorems 3.2 and 3.3. Shrinkage strictly improves the asymptotic risk, and the improvement can be especially strong in high-dimensional cases. This holds true for other n and r values.

For the LASSO and adaptive LASSO estimators, we use the local quadratic approximation algorithm of Ulbricht (2010) for finding the entire solution path. We also use a k-fold cross validation procedure in order to choose the value of the penalization parameter τ that achieves the lowest BIC score. Table 1 shows that when $r < 6$, the LASSO and adaptive LASSO estimators are better than the PSE while it performs well for $r \geq 6$. The PSE is preferable in the presence of many inactive predictors in the model.

Table 1. Simulated relative efficiency of RE, SE, PSE, LASSO, and adaptive LASSO with respect to $\hat{\boldsymbol{\theta}}$ when the restricted parameter space $\boldsymbol{H\theta} = \boldsymbol{h}$ is correct ($\Delta = 0$).

		n=150		
Method	$p_2 = 3$	$p_2 = 6$	$p_2 = 11$	$p_2 = 14$
RE	3.58	3.70	4.08	4.11
PT	2.31	2.62	3.11	3.02
SE	1.39	1.93	2.76	2.87
PSE	1.47	2.23	2.87	3.07
LASSO	1.48	1.86	2.51	2.75
ALASSO	1.51	1.95	2.61	2.83
		n=350		
RE	2.71	2.30	2.77	3.17
PT	2.14	2.04	2.11	2.59
SE	1.27	1.62	2.06	2.43
PSE	1.41	1.79	2.23	2.67
LASSO	1.43	1.53	1.63	1.67
ALASSO	1.49	1.60	1.72	1.81

5. Application to Real Life Data

The data are from the National Medical Expenditure Survey (NMES) conducted in 1987 and 1988 to provide a comprehensive picture of how Americans use and pay for health services. It includes 4406 respondents aged 66 or older and covered by Medicare program. Details of the data can be found in Deb & Trivedi (1997). In this example, the number of hospital stays (hosp) is used as the response variable and the covariates are self-perceived health status (poorhlth and exclhlth), number of chronic conditions (numchron), gender (male), age, number of years of education (school), family income (faminc), and private insurance indicator (privins). Approximately 80% of the responses are equal to zero, which corresponds to patients with zero day stay in the hospital. We first look at the count portion of the data, which refers to the respondents who have stayed at least one day in the hospital. On the other hand, if we look at the logistic portion (i.e. inflation part) of the data, which predict whether outcome is positive or zero. All covariates are included in both portions of the model to predict the number of hospital stays. To assess the effect of covariates on the hosp on each portion, ZIPR model is fitted. The AIC and BIC criteria for the mixture model (Wang et al., 1996) show that poorhlth, exclhlth and numchron are the important factors for the Poisson part and poorhlth, numchron, age, and male are the important factors for the logistic part of the ZIPR model. So the other six covariates in the Poisson part and five covariates in the logistic part are not significantly related to the response hosp. In this example, our hypothesis is $\boldsymbol{H\theta} = \boldsymbol{0}$ where $\boldsymbol{H}$ is a 7×18 matrix and $\boldsymbol{\theta}$ is a 18×1 vector of covatiates.

We apply the bootstrap method (see, Jung et al., 2005) to examine the performance of the proposed estimation strategies for estimating the coefficients of the other seven covariates. We draw 1000 bootstrap samples of size $n = 2000$ by drawing 2000 rows with replacement from the data matrix (y_i, x_{ij}) and compute the point estimates, standard errors, and relative efficiencies of the proposed estimators. Here, the empirical distribution function $\hat{F}$ based on original 4406 individuals is regarded as the true distribution, and the coefficients of the ZIPR closest to $\hat{F}$ are regarded as true parameter values. The results given in Table 2 reveal that the restricted, pretest, shrinkage, positive shrinkage, LASSO, and adaptive LASSO estimators are superior to the maximum likelihood estimator, which is in agreement with our simulation results.

Table 2. Estimate (first row) and standard error (second row) for poorhlth (β_1), exclhlth (β_2), numchron (β_3) of the Poisson part and poorhlth (γ_1), numchron (γ_2), age (γ_3), and male (γ_4) of logistic part on the number of hospital stays. The SRE column gives the relative efficiency based on bootstrap simulation of the estimators with respect to UMLE.

Estimators	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4	SRE
UMLE	-0.99	-0.313	0.123	-0.637	-0.302	-0.446	-0.381	1.000
	0.502	0.125	0.048	0.256	0.095	0.153	0.246	
RMLE	-0.715	-0.335	0.118	-0.555	-0.297	-0.375	-0.286	1.444
	0.186	0.127	0.042	0.230	0.079	0.086	0.107	
PT	-0.882	-0.321	0.122	-0.609	-0.301	-0.421	-0.354	1.194
	0.448	0.126	0.046	0.255	0.090	0.131	0.214	
SE	-0.861	-0.323	0.121	-0.601	-0.300	-0.414	-0.342	1.226
	0.334	0.124	0.045	0.243	0.086	0.117	0.182	
PSE	-0.863	-0.322	0.121	-0.602	-0.300	-0.415	-0.343	1.226
	0.332	0.124	0.044	0.244	0.086	0.117	0.182	
LASSO	-0.543	-0.190	0.121	-0.435	-0.254	-0.069	-0.417	1.178
	0.267	0.128	0.057	0.291	0.106	0.089	0.099	
ALASSO	-0.473	-0.197	0.121	-0.439	-0.257	-0.068	-0.445	1.221
	0.231	0.136	0.058	0.248	0.106	0.073	0.098	

6. Conclusion

In this paper, we have introduced the pretest, shrinkage, LASSO, and ALASSO estimators for zero-inflated Poisson regression model when it is suspected that some of the regression coefficients may be restricted to a subspace. We have presented ADB and ADR expressions of the pretest and shrinkage estimators. We conducted a Monte Carlo simulation experiment to examine the performance of the proposed estimators which showed that the RMLE outperforms the usual UMLE at or near the restriction. However, as we deviate from the the restriction, risk of the RMLE becomes unbounded. And near the restriction, the risk of the pretest estimator is less than that of the UMLE. As Δ, the distance between the simulated and restricted models increases, the risk of the pretest estimator crosses the risk of the UMLE, reaching a maximum, and then decreasing monotonically to the risk of the ULME. Furthermore, the shrinkage estimators with data based weights perform well if the restriction is true. In fact, the shrinkage estimators outperform the UMLE in the entire parameter space when the number of inactive parameter is greater than two. In addition, the performance of the shrinkage estimators improve relative to the UMLE when the number of inactive covariates increases. The penalty estimators are preferable when there are fewer inactive covariates. Thus in presence of many inactive covariates, the shrinkage estimators are attractive alternatives to penalty estimators.

Finally, we applied the proposed strategies to a real life data set to evaluate the relative performance of the proposed estimators. The results are consistent with our analytical and simulated findings. The theoretical and numerical results of zero-inflated Poisson can be extended to the entire class of generalized linear models. We are currently working on this project.

In terms of recommendations, the shrinkage estimators should be used instead of the UMLE when the restriction on regression coefficients agrees with the data. If we are uncertain of the quality of auxiliary information about the covariates, we can still use the shrinkage estimation because it offers a lower or at best an equal risk relative to the UMLE over the entire parameter space.

Acknowledgements

The authors thank the Editor and the referees for their helpful comments which has improved the earlier version of the manuscript. The research of Shakhawat Hossain was supported by the Natural Sciences and the Engineering Research Council of Canada grant and that of Hatem A. Howlader was supposted by a grant from the University of Winnipeg.

References

Ahmed, S. E., Hussein, A. A., & Sen, P. K. (2006). Risk comparison of some shrinkage M-estimators in linear models. *Journal of Nonparametric Statistics, 18* (4-6), 401-415. http://dx.doi.org/10.1080/10485250601046752

Ahmed, S. E., & Fallahpour, S. (2012). Shrinkage estimation strategy in quasi-likelihood models. *Statistics and Probability Letters, 82*, 2170-2179. http://dx.doi.org/10.1016/j.spl.2012.08.001

Deb, P., & Trivedi, P. K. (1997). Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Economics, 12*(3), 313-336. http://dx.doi.org/10.1002/(SICI)1099-1255(199705)12:3¡313::AID-JAE440¿3.0.CO;2-G

Dietz, E., & Bohning, D. (2000). On estimation of the Poisson parameter in zero-inflated Poisson models. *Computational Statistics and Data Analysis, 34*(4), 441-459. http://dx.doi.org/10.1016/S0167-9473(99)00111-5

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2006). Least Angle Regression. *Annals of Statistics, 32*, 407-499.

Feng, J., & Zhu, Z. (2012). Semiparametric analysis of longitudinal zero-inflated count data. *Journal of Multivariate Analysis, 102*, 61-72. http://dx.doi.org/10.1016/j.jmva.2010.08.001

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*, 1-22.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics, 56*, 1030-1039. http://dx.doi.org/10.1111/j.0006-341X.2000.01030.x

Hinde, J. P., & Demetrio, C. G. B. (1998). Over-dispersion: Models and estimation. *Computational Statistics and Data Analysis, 27*, 151-170. http://dx.doi.org/10.1016/S0167-9473(98)00007-3

Hossain, S., Doksum, K. A., & Ahmed, S. E. (2009). Positive-part shrinkage and Absolute Penalty Estimators in Partially Linear Models. *Linear Algebra and its Applications, 430*(3), 2749-2761. http://dx.doi.org/10.1016/j.laa.2008.12.015

Hossain, S., Ahmed, S. E., & Howlader, H. A. (2012). Model selection and parameter estimation of a multinomial logistic regression model. *Journal of Statistical Computation and Simulation, 84*(7), 1412-1426. http://dx.doi.org/10.1080/00949655.2012.746347

Hossain, S., & Ahmed, S. E. (2012). Shrinkage and Penalty Estimators of a Poisson Regression Model. *Australian and New Zealand Journal of Statistics, 54*(3), 359-373. http://dx.doi.org/10.1111/j.1467-842X.2012.00679.x

Jackman, S. (2012). pscl: Political Science Computational Laboratory, Stanford University. *R package version 1.4.9.* http://CRAN.R-project.org/package=pscl

Jansakul, N., & Hinde, J. P. (2002). Score tests for zero-inflated Poisson models. *Computational Statistics and Data Analysis, 40*, 75-96. http://dx.doi.org/10.1016/S0167-9473(01)00104-9

Judge, G.G., & Bock, M. E. (1978). *The Statistical Implications of pretest and Stein-Rule Estimators in Econometies*. Amsterdam: North Holland Publishing Company.

Jung, B. C., Jhun, M., & Lee, J. W. (2005). Bootstrap Tests for Overdispersion in a Zero-Inflated Poisson Regression Model. *Biometrics, 61*, 626-629. http://dx.doi.org/10.1111/j.1541-0420.2005.00368.x

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics, 34*, 1-14. http://dx.doi.org/10.2307/1269547

Liang, H., & Song, W. (2009). Improved estimation in multiple linear regression models with measurement error and general constraint. *Journal of Multivariate Analysis, 100*, 726-741. http://dx.doi.org/10.1016/j.jmva.2008.08.003

McCullagh, P. (1983). Quasi-likelihood functions. *Annals of Statistics, 11*, 59-67. http://dx.doi.org/10.1214/aos/1176346056

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics, 33*(3), 341-365. http://dx.doi.org/10.1016/0304-4076(86)90002-3

Nkurunziza, S. (2013). Extension of some important identities in shrinkage-pretest strategies. *Metrika, 76*(7), 937-947. http://dx.doi.org/10.1007/s00184-012-0425-5

Park, M. Y., & Hastie, T. (2007). An L_1 regularization-path algorithm for generalized linear models. *Journal of the Royal Statistical Society: Series B, 69*, 659-677. http://dx.doi.org/10.1111/j.1467-9868.2007.00607.x

Ridout, M. S., Demetrio, C. G. B., & Hinde, J. P. (1998). *Models for count data with many zeros*. In: International Biometric Conference. Cape Town, pp. 179-190.

Ridout, M. S., Hind, J. P., & Demetrio, C. G. B. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics, 57*, 219-223. http://dx.doi.org/10.1111/j.0006-341X.2001.00219.x

Sapra, S. K. (2003). Pre-test estimation in Poisson regression model. *Applied Economics Letters, 10*, 541-543. http://dx.doi.org/10.1080/1350485032000100215

Tibshirani, R. (1996). Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society: Series B, 58*(7), 267-288.

van den Broek, J. (1995). A score test for zero inflation in a Poisson distribution. *Biometrics, 51*, 738-743. http://dx.doi.org/10.2307/2532959

Wang, P., Puterman, M. L., Cockburn, I., & Le, N. (1996). Mixed Poisson regression models with covariate dependent rates. *Biometrics, 52*, 381-400. http://dx.doi.org/10.2307/2532881

Zeng, H., Wei, Y., Zhao, Y., Liu, L., Zhang, R., Gou, J., Huang, S., & Chen, F. (2014). Variable selection approach for zero-inflated count data via adaptive LASSO. *Journal of Applied Statistics, 41*(4), 879-894. http://dx.doi.org/10.1080/02664763.2013.858672

Zou, H. (2006). The Adaptive LASSO and its oracle properties. *Journal of the American Statistical Association, 101*, 1418-1429. http://dx.doi.org/10.1198/016214506000000735

International Journal of Statistics and Probability; Vol. 4, No. 4; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Inferences for a Two-parameter Lifetime Distribution with Bathtub Shaped Hazard Based on Censored Data

Ammar M. Sarhan[1] & Joseph Apaloo[2]

[1] Department of Mathematics and Statistics, Dalhousie University, Canada

[2] Department of Mathematics, Statistics & Computer Sciences, St. Francis Xavier University, Canada

Correspondence: Ammar M. Sarhan, Department of Mathematics and Statistics, Dalhousie University, Halifax NS B3H 4R2, Canada. Tel: 1-902-266-6423. E-mail: asarhan@mathstat.dal.ca

Received: April 20 Accepted: May 4, 2015 Online Published: October 9, 2015

doi:10.5539/ijsp.v4n4p77 URL: http://dx.doi.org/10.5539/ijsp.v4n4p77

Abstract

We consider statistical inference of the unknown parameters of a two-parameter bathtub-shaped distribution (Chen, 2000) [Stat. & Prob. Letters 49 (2000) 155-161]. The inference will be conducted for Type-II censored and progressively Type-II censored data using the maximum likelihood and Bayes techniques. There are no explicit expressions for the estimators of the parameters. In the case of the maximum likelihood estimator (MLE), we propose a simple fixed point algorithm to compute the MLE and construct different confidence intervals and confidence regions of the unknown parameters. Bayes analyses of the unknown parameters are also discussed under fairly general priors for the unknown parameters. We propose to use the Markov Chain Monte Carlo (MCMC) and simulation-based technique to compute the Bayes estimates and the two-sided Bayesian probability intervals of the parameters. Also, we use the rejection sampling algorithm to produce the exact Bayes estimates. The methods developed will be applied in the analyses of two real data sets and a simulated data set. A Monte Carlo simulation is used to compare the results from the MLE and Bayes techniques.

Keywords: maximum likelihood estimator, MCMC, rejection sampling, simulation-based.

1. Introduction

The cumulative distribution function (cdf) of the Weibull distribution with scale parameter α and shape parameter β denoted by $W(\alpha,\beta)$ is

$$F_W(x;\alpha,\beta) = 1 - e^{-\alpha x^{\beta}},\ x \geq 0, \alpha,\beta > 0. \tag{1}$$

The hazard rate function of $W(\alpha,\beta)$, $h_W(x;\alpha,\beta) = \alpha\beta x^{\beta-1}$, is increasing in x when $\beta > 1$, decreasing when $\beta < 1$ and constant when $\beta = 1$.

The two parameter Weibull distribution is a very popular distribution for modeling phenomenon with monotone hazard rates. Its negatively and positively skewed density makes it an initial choice for modeling monotone hazard rates. However, Weibull distribution does not provide a reasonable fit in modeling phenomenon with non-monotone hazard rate such as the bathtub-shaped hazard rate. A distribution with a bathtub-shaped hazard rate provides an appropriate conceptual model for some electronic and mechanical products as well as the lifetime of humans.

There is extensive literature on parametric probability distributions with bathtub-shaped hazard rate function (e.g., Smith and Bain (1975), Leemis (1986), Gaver and Acar (1979), Hjorth (1980), and Mudhilkar and Srivastava (1993)). Among several extensions of the Weibull distribution, with 4 parameters, are the exponentiated generalized exponential linear distribution (Sarhan et. al, 2013) and the exponentiated modified Weibull extension distribution (Sarhan and Apaloo, 2013).

Chen (2000) revisited a two parameter distribution with the following survival function (sf)

$$S(x;\alpha,\beta) = e^{\alpha\left(1-e^{x^{\beta}}\right)},\ x > 0, \tag{2}$$

where $\alpha > 0$ is a parameter which does not affect the shape of the failure rate function and $\beta > 0$ is the shape parameter. The corresponding probability density function (pdf) is

$$f(x;\alpha,\beta) = \alpha\beta x^{\beta-1} e^{x^{\beta}} e^{\alpha\left(1-e^{x^{\beta}}\right)},\ x > 0. \tag{3}$$

The corresponding hazard rate function (hrf) of this distribution is

$$h(x;\alpha,\beta) = \alpha\beta x^{\beta-1}e^{x^{\beta}},\ x > 0. \tag{4}$$

The hrf can be either (1) an exponentially increasing function when $\beta \geq 1$; or (2) of bathtub-shape when $\beta < 1$. For simplicity, we will refer to this distribution as TPBT(α,β). It has been observed in Nadarajah and Kotz (2007) and Pham and Lai (2007), that TPBT(α,β) is a special case of the distribution considered by Gurvich et al. (1997) and by Haynatzki et al. (2000). However, TPBT(α,β) has an advantage over all bathtub hazard shaped distributions since it has only two parameters.

The TPBT(α,β) distribution has been considered for a type-II censored data (Chen, 2000), type-II right censored data (Wu et al., 2004), for a progressively type-II censored data (Wu, 2008). In Chen (2000), exact onfidence interval for β and exact joint confidence regions for (α,β) were presented while Wu et al. (2004) presents statistical inference about the shape parameter. Sarhan *et al.* (2012) derived both the maximum likelihood estimates (MLE) and Bayes estimates (BE) of the two unknown parameters using a complete sample. They applied the simulation-based (SB) method, the Monte Carlo integration (MCI) technique and Lindley approximation (LA) method to approximate the Bayes estimates of the parameters and reported that the SB method was the best among the three methods used.

The aim of this paper is to discuss Bayes inferences of the TPBT distribution's parameters using different techniques when the data are complete, cenceored and progressively censored. We use Markov-Chain Monte-Carlo (MCMC), rejection sampling (RS) algorithms and compare them with the SB algorithm.

The rest of the paper is organized as follows. In section 2 we present the censored data scheme and the maximum likelihood estimation procedure. Confidence interval estimations are discussed in section 3. Bayes procedure is discussed in section 4. Two approximation techniques (the RS and MCMC) are presented in section 5. Progressively Type-II censored data is used in section 6. Two real and a simulated data sets are analyzed in section 7. A large simulation study is provided in section 8 to compare the maximum likelihood technique with Bayesian techniques. Concluding remarks are made in section 9.

2. Maximum Likelihood Estimation

In Type-II censored data, it is assumed that n independent and identical units are put on the life test at the same time. The life test is terminated after a predetermined number of failures m results. For unit i, $i = 1, 2, \cdots, n$, a pair of two quantities $(X_i,\ \delta_i)$ is observed, where X_i represents the testing time of unit i and δ_i is a binary variable that takes the value of 1 when unit i has failed at time X_i, or 0 if it is tested up to time X_i without failure (X_i is a censored time).

In this section we use the maximum likelihood method to estimate the two unknown parameters α and β, using Type-II censored data. Suppose (X_1,δ_1), (X_2,δ_2), ..., $(X_n,,\delta_n)$ is a random sample from TPBT(α,β), then the likelihood function of the observed data is

$$L = (\alpha\beta)^m \left[\prod_{i=1}^{n} x_i^{\delta_i}\right]^{\beta-1} e^{\sum_{i=1}^{n}\left\{\delta_i x_i^{\beta}+\alpha\left(1-e^{x_i^{\beta}}\right)\right\}}, \tag{5}$$

where $m = \sum_{i=1}^{n}\delta_i$. The log-likelihood function becomes

$$\mathcal{L} = m(\ln\alpha + \ln\beta) + (\beta-1)\sum_{i=1}^{n}\delta_i \ln x_i + \sum_{i=1}^{n}\left\{\delta_i x_i^{\beta} + \alpha\left(1-e^{x_i^{\beta}}\right)\right\}. \tag{6}$$

Taking derivatives with respect to α and β of (6), we obtain

$$\frac{\partial\mathcal{L}}{\partial\alpha} = \frac{m}{\alpha} + n - \sum_{i=1}^{n} e^{x_i^{\beta}}, \tag{7}$$

$$\frac{\partial\mathcal{L}}{\partial\beta} = \frac{m}{\beta} + \sum_{i=1}^{n}\delta_i \ln x_i + \sum_{i=1}^{n}\delta_i x_i^{\beta}\ln x_i - \alpha\sum_{i=1}^{n} e^{x_i^{\beta}} x_i^{\beta}\ln x_i. \tag{8}$$

The second partial derivatives of $\mathcal{L}$ with respect to α and β are

$$\begin{aligned}
\mathcal{L}_{11} &= \frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = -\frac{m}{\alpha^2}\,, \\
\mathcal{L}_{12} &= \frac{\partial^2 \mathcal{L}}{\partial \alpha \beta} = -\sum_{i=1}^{n} x_i^{\beta}\, e^{x_i^{\beta}}\, \ln x_i = \mathcal{L}_{21}\,, \\
\mathcal{L}_{22} &= \frac{\partial^2 \mathcal{L}}{\partial \beta^2} = -\frac{m}{\beta^2} + \sum_{i=1}^{n} \delta_i\, x_i^{\beta}\, (\ln x_i)^2 - \alpha \sum_{i=1}^{n} x_i^{\beta} \left(1 + x_i^{\beta}\right) (\ln x_i)^2\, e^{x_i^{\beta}}\,.
\end{aligned}$$

The information function is a two-by-two symmetric matrix

$$\mathfrak{I}(\alpha,\beta) = \begin{bmatrix} \mathfrak{I}_{11}(\alpha,\beta) & \mathfrak{I}_{12}(\alpha,\beta) \\ \mathfrak{I}_{21}(\alpha,\beta) & \mathfrak{I}_{22}(\alpha,\beta) \end{bmatrix} = -\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} \\ \mathcal{L}_{21} & \mathcal{L}_{22} \end{bmatrix}. \tag{9}$$

For a relative maximum of $\mathcal{L}$, which occurs at the MLE of the parameters $\hat{\alpha}$ and $\hat{\beta}$, the matrix $\mathfrak{I}(\hat{\alpha},\hat{\beta})$ must be positive definite.

Setting the right hand side of (7) to zero and solving for α, we get

$$\hat{\alpha}(\beta) = \frac{m}{\sum_{i=1}^{n} e^{x_i^{\beta}} - n}\,, \tag{10}$$

Substituting (10) into the right side of (8) and setting it equal to zero, we get the following non-linear equation in β

$$\frac{m}{\beta} + \sum_{i=1}^{n} \delta_i\, \ln x_i + \sum_{i=1}^{n} \delta_i\, x_i^{\beta} \ln x_i - \frac{m \sum_{i=1}^{n} e^{x_i^{\beta}} x_i^{\beta} \ln x_i}{\sum_{i=1}^{n} e^{x_i^{\beta}} - n} = 0\,. \tag{11}$$

The MLE of β, $\hat{\beta}$, is the solution of (11) in β. A closed-form solution of (11) does not exist, so a numerical technique, e.g Newton-Raphson method, should be used to find $\hat{\beta}$ for any given data set. Once we get $\hat{\beta}$, we can use (10) to get $\hat{\alpha} = \hat{\alpha}(\hat{\beta})$.

3. Confidence Intervals (CIs)

We can use different techniques to approximate $(1-\vartheta)100\%$ confidence intervals of the two parameters α and β. In the following, we describe two of such techniques.

3.1 Likelihood Intervals

A $100p\%$ likelihood interval (LI), $p \in (0,1)$, for a parameter θ is the range of all values of θ for which the relative likelihood function, $R(\theta) = \frac{L(\theta)}{L(\hat{\theta})}$, is greater than or equal to p, for more details we refer to Kalbfleisch (1985). For simplicity, we can take the natural logarithm of $R(\theta)$, therefore the $100p\%$ LI of θ becomes the range of all values of θ which satisfy that $r(\theta) \geq \ln p$. Here, $r(\theta)$ is the log-relative likelihood function, given by $r(\theta) = \mathcal{L}(\theta) - \mathcal{L}(\hat{\theta})$. To determine the $100p\%$ LI for θ, we can either use the the graph of $r(\theta)$ against θ or use the Newton-Raphson Method to get the bounds of the LI by solving $r(\theta) - \ln p = 0$.
When θ is a vector of two unknown parameters, say $\theta = (\alpha,\beta)'$, as in the case studied here, the $100p\%$ likelihood region is the set of parameter values (α,β) such that $r(\alpha,\beta) \geq \ln p$, where $r(\alpha,\beta)$ is the joint log-relative likelihood function of α and β, $r(\alpha,\beta)$, given by

$$r(\alpha,\beta) = \mathcal{L}(\alpha,\beta) - \mathcal{L}(\hat{\alpha},\hat{\beta}). \tag{12}$$

Substituting from (6) into (12), we get

$$r(\alpha,\beta) = m(\ln\alpha + \ln\beta) + (\beta-1)K_0 + K_1(\beta) - \alpha\, K_2(\beta) - \mathcal{L}(\hat{\alpha},\hat{\beta}), \tag{13}$$

where $K_0 = \sum_{i=1}^{n} \delta_i \ln x_i$, $K_1(\beta) = \sum_{i=1}^{n} \delta_i\, x_i^{\beta}$ and $K_2(\beta) = \sum_{i=1}^{n} e^{x_i^{\beta}} - n$. The $100p\%$ likelihood contour is boundary of $100p\%$ likelihood region, which is formed by the curve $r(\alpha,\beta) = \ln p$. The $100p\%$ likelihood region for (α,β) is an approximate $100(1-\vartheta)\%$ confidence region when $p = e^{-\chi_1^2(\vartheta)/2}$, where $\chi_1^2(\vartheta)$ is the upper ϑ quantile of the chi square distribution with one degree of freedom.

Another way to get the $100p\%$ LI is to use the maximum log-relative likelihood function of β, which is given by

$$r_{\max}(\beta) = m[\ln(m\beta) - 1] - \ln(K_2(\beta)) + (\beta-1)K_0 + K_1(\beta) - \mathcal{L}(\hat{\alpha},\hat{\beta}). \tag{14}$$

A $100p\%$ LI for β, say $[\beta_1, \beta_2]$, is the set of all β values such that

$$r_{\max}(\beta) \geq \ln p\,. \tag{15}$$

Inequality (15) has no closed form solution in β, therefore a numerical method, such as Newton-Raphson, should be used to get $[\beta_1, \beta_2]$. Using $[\beta_1, \beta_2]$, we can derive a $100p\%$ LI for α as

$$\frac{m}{\sum_{i=1}^{n} e^{x_i^{\beta_2}} - n} \leq \alpha \leq \frac{m}{\sum_{i=1}^{n} e^{x_i^{\beta_1}} - n}\,. \tag{16}$$

3.2 Large-sample Intervals

The MLE of the parameters α and β are asymptotically normally distributed with means equal to the true values of α and β and variances given by the inverse of the information matrix. In particular,

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \sim N_2\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \hat{\mathfrak{I}}^{-1}\right), \tag{17}$$

where $\hat{\mathfrak{I}}^{-1}$ is the inverse of $\mathfrak{I}(\hat{\alpha},\beta)$, with main diagonal elements $\hat{\mathfrak{I}}^{11}$ and $\hat{\mathfrak{I}}^{22}$ given by

$$\hat{\mathfrak{I}}^{11} = \frac{\hat{\mathfrak{I}}_{22}}{\hat{\mathfrak{I}}_{11}\hat{\mathfrak{I}}_{22} - \hat{\mathfrak{I}}_{12}^2} \text{ and } \hat{\mathfrak{I}}^{22} = \frac{\hat{\mathfrak{I}}_{11}}{\hat{\mathfrak{I}}_{11}\hat{\mathfrak{I}}_{22} - \hat{\mathfrak{I}}_{12}^2}\,.$$

Using (17), large-sample $(1-\vartheta)100\%$ confidence intervals for α and β are

$$\hat{\alpha} \pm Z_{\vartheta/2}\sqrt{\hat{\mathfrak{I}}^{11}} \quad \text{and} \quad \hat{\beta} \pm Z_{\vartheta/2}\sqrt{\hat{\mathfrak{I}}^{22}}\,,$$

where $Z_{\vartheta/2}$ is the upper $\vartheta/2$ quantile of the standard normal distribution.

4. Bayes Inferences

To make Bayes inferences about the parameters α and β, we assume that α and β are independent random variables having gamma prior distributions with parameters (a_1, a_2) and (b_1, b_2), respectively. Using the results in Sarhan et al. (2012), we can get the posterior joint pdf of (α,β), the marginal posterior distributions of α and β, the Bayes point estimates of α and β and the corresponding minimum Bayes risks, and the two-sided Bayesian probability intervals for α and β, by replacing $P_x = \prod_{i=1}^{n} x_i$ and $K_1(\beta) = \sum_{i=1}^{n} x_i^{\beta}$ by $P_x = \prod_{i=1}^{n} x_i^{\delta_i}$ and $K_1(\beta) = \sum_{i=1}^{n} \delta_i x_i^{\beta}$, respectively, and using the same $K_2(\beta) = \sum_{i=1}^{n}\left(e^{x_i^{\beta}} - 1\right)$. The posterior density of (α,β) is

$$\pi(\alpha,\beta|data) = \frac{1}{I_{(0,0)}}\,\alpha^{m+a_1-1}\beta^{m+b_1-1}\,e^{-[a_2+K_2(\beta)]\alpha - b_2\beta + K_1(\beta) + (\beta-1)\sum_{i=1}^{n}\delta_i \ln x_i}\,, \tag{18}$$

where

$$I_{(0,0)} \;=\; \Gamma(m+a_1)\int_0^{\infty} \frac{\beta^{m+b_1-1}}{[a_2+K_2(\beta)]^{m+a_1}}\,e^{-b_2\beta + K_1(\beta) + (\beta-1)\sum_{i=1}^{n}\delta_i \ln x_i}\,d\beta\,, \tag{19}$$

$K_1(\beta) = \sum_{i=1}^{n} \delta_i\, x_i^{\beta}$ and $K_2(\beta) = -n + \sum_{i=1}^{n} e^{x_i^{\beta}}$.

As it is stated in Sarhan et al. (2012), all Bayes results have no analytic solutions in the general case. Therefore, numerical approximations and/or simulation techniques should be used.

5. Approximation to Bayes Estimates

Sarhan et al. (2012) used (i) Simulation-based method (SB); (ii) Monte-Carlo integration (MCI) method; and (iii) Lindley approximation method (LA), to make Bayes inferences about the two parameters and they concluded, based on a simulation study, that SB method provides better estimates than the other two methods. In this paper we use rejection sampling (RS) algorithm and Markov Chain Monte Carlo (MCMC) techniques to perform Bayesian inferences about the model parameters and compare it with SB method.

5.1 Rejection Sampling (RS)

The rejection sampling technique is used to produce simulated independent samples from a given density function. For more details about the RS, we refer the readers to Albert (2009) and Gelman et al. (2014). Suppose we want

to sample from the density $p(\theta|\text{data})$, which is not an easy function to simulate from. The rejection sampling requires a positive function $g(\theta)$ defined for all θ for which $p(\theta|\text{data}) > 0$ such that: (1) we can draw from a probability density function proportional to g; (2) it is not required that $g(\theta)$ be a density function, but must have a finite integral; (3) $p(\theta|\text{data}) < Cg(\theta)$ for all θ, where $C > 1$ is an appropriate bound on $\frac{p(\theta|\text{data})}{g(\theta)}$. The functions $p(\theta|\text{data})$ and $g(\theta)$ are called the target and proposal functions, respectively. In our case here, the target function is the posterior density (18). The main task in the rejection sampling is to find a suitable proposal density $g(\theta)$ and constant value C that satisfy the above restriction.
The following is the RS algorithm to simulate draws from (18):

1. Simulate $\theta = (\alpha, \beta)$ from the proposal $g(\theta)$.
2. Simulate U from a uniform distribution on the unit interval.
3. If $U \leq \frac{\pi(\theta|data)}{Cg(\theta)}$, then accept θ as a draw from (18).
4. Repeat steps 1-3 N times.

The main advantage of the RS algorithm is the accepted θ has the correct posterior distribution.

The main problem with this process is that C is generally large in high-dimensional spaces and since $P(\text{accept}) \propto \frac{1}{C}$, many samples will get rejected.

5.2 Markov Chain Monte Carlo (MCMC)

Since early 1990s, Markov Chain Monte Carlo (MCMC) techniques have been used extensively because of their generality and flexibility along with the massive development of computing facilities. It has become one of the main computational tools in modern Bayesian statistical inference. Metropolis et al. (1953) developed a simple version of MCMC, known as the Metropolis-Hastings algorithm (MHA). A generalization of the MHA was proposed by Hastings (1970). A comprehensive theoretical exposition of this algorithm is given by Tierney (1994) and an excellent tutorial of this topic is provided by Chib and Greenberg (1995).

MCMC methods enable quantitative scientists to use highly complicated models and estimate the corresponding posterior distributions with accuracy. For extensive details of the use of MCMC methods, we refer readers to Gilks et al. (1996) and Chen et al. (2000). In this paper, we use the Metropolis-Hastings algorithm (MHA) to generate samples from the posterior distribution $\pi(\theta|data)$ in (18).

The MHA can be described by the following iterative steps:

1. Set initial values $\theta^{(0)} = \left(\alpha^{(0)}, \beta^{(0)}\right)'$.
2. For $t = 1, \cdots, T$ repeat the following steps:
   i. Set $\theta = \theta^{(t-1)}$.
   ii. Generate a new candidate θ^* from a proposal distribution $q(\theta^*|\theta)$.
   iii. Calculate
   $$\kappa = \min\left\{1, \frac{\pi(\theta^*|data)}{\pi(\theta|data)}\right\}$$
   iv. Set $\theta^{(t)} = \theta^*$ with probability κ; or otherwise set $\theta^{(t)} = \theta$.

The MHA converges to its equilibrium distribution regardless of the proposal distribution q. Neverthless, in practice, choice of the proposal is important since a poor choice considerably delays the convergence towards the equilibrium distribution. An efficient proposal, for the underlying model, is the bivariate normal with mean of the MLE of θ and covariance matrix of the Fisher-information matrix evaluated at the MLE of θ (Lecam, 1986).

6. Progressive Type-II Censored Data

In this section, we use progressive type-II censored data from TPBT distribution to make inference about the distribution parameters using Bayes technique. In this data, it is assumed that n independent and identical units are placed on a life test simultaneously with m failures to be observed. When the first failure occurs, r_1 survived units are randomly removed from the test. At the second failure, r_2 survived units are randomly removed. The test

stops at the occurrence of m-th failure and the remaining $r_m = n - \sum_{i=1}^{m-1} r_i - m$ survived units are all removed. The likelihood function in this case is

$$L = \alpha^m \beta^m \left(\prod_{i=1}^{m} x_i \right)^{\beta-1} \exp\left\{ \sum_{i=1}^{m} x_i^{\beta} + \alpha \sum_{i=1}^{m} (r_i + 1) \left(1 - e^{x_i^{\beta}} \right) \right\}.$$

The log-likelihood function is

$$\mathcal{L} = m \ln \alpha + m \ln \beta + (\beta - 1) \sum_{i=1}^{m} \ln x_i + \sum_{i=1}^{m} x_i^{\beta} + \alpha \sum_{i=1}^{m} (r_i + 1) \left(1 - e^{x_i^{\beta}} \right).$$

Wu (2008) obtained the MLE of the parameters α and β and discussed their confidence intervals and joint confidence region using a progressively type-II censored sample. Wu (2008) addressed that the exact confidence interval and the exact confidence region are obtained, but (*i*) numerical approximation is used to get the confidence interval of β and no confidence interval of α was discussed, and (*ii*) the confidence limits for β, which are needed to get the confidence region for (α, β), are obtained numerically.

In this paper we discuss the likelihood intervals for both α and β, the likelihood region for (α, β), and Bayes analysis of the two parameters α and β using progressively type-II censored sample. Using all results in the previous sections, we can get the corresponding results for maximum likelihood (such as the likelihood intervals and likelihood region) and Bayes techniques (point estimate and two-sided Bayesian intervals) using a progressively type-II censored sample by replacing $P_x = \prod_{i=1}^{n} x_i^{\delta_i}$, $K_1(\beta) = \sum_{i=1}^{n} \delta_i x_i^{\beta}$, and $K_2(\beta) = \sum_{i=1}^{n} \left(e^{x_i^{\beta}} - 1 \right)$ by $P_x = \prod_{i=1}^{m} x_i$, $K_1(\beta) = \sum_{i=1}^{m} x_i^{\beta}$, and $K_2(\beta) = \sum_{i=1}^{m} (r_i + 1) \left(e^{x_i^{\beta}} - 1 \right)$, respectively. Notice that $K_0 = \ln P_x$.

7. Applications

In this section, two applications are discussed and one illustrative example. All the computations were done using Matlab 7.11.0 (R2010) programming language, except for RS we used R programming language. In all calculations, the values of M and T for SB and MCMC are 20000 and $N = 10000$ for RS, also since we do not have any prior information, we assume $a_1 = a_2 = b_1 = b_2 = 0.001$. Although these values of the prior parameters imply improper priors on α and β, but the corresponding joint posterior is proper.

Application 1: Serum data

In this application, the data set refers to the serum-reversal time (days) of 148 children contaminated with HIV from vertical transmission at the university hospital of the Ribeiro Preto School of Medicine (Hospital das Clnicas da Faculdade de Medicina de Ribeiro Preto) from 1986 to 2001 (Silva, 2004) . For more details, we refer the readers to Perdon (2006). We start with the TPBT(α, β) and W(α, β) distributions for fitting and classical inferences. Then, we will focus on the Bayesian inference of the TPBT(α, β) model.

It is assumed that the lifetimes are independent and identically distributed, and also independent from the censoring mechanism. Also, we assumed that the lifetimes follow either Weibull or TPBT model. In order to assess which model is more appropriate to fit this data set, we plot the scaled TTT-transform (Figure 1.a), the parametric hazard function (Figure 1.b), the empirical and estimated survival functions of the TPBT(α, β) and W(α, β) distributions (Figure 2). Figure 1.a shows that the scaled TTT-plot for the data set has first a convex shape and then a concave shape which indicates a bathtub-shaped hrf. Figure 1.b shows that the TPBT model has a bathtub hrf while Weibul has an increasing hrf. Therefore, the TPBT model would be an appropriate model for the fitting of this data. The MLEs of the model parameters and the corresponding values of the $\mathcal{L}$ (the log-likelihood function), AIC (Akaike Information Criterion), K-S (Kolmogorov-Smirnov) test statistic and the corresponding p-value are given in Table 1. These results show that the AIC of the TPBT is smaller than that for Weibull which indicates that the TPBT model is more appropriate to fit this data than the Weibull. Furthermore, based on the K-S test the Weibull model is rejected at any significance level greater than 0.025 while the TPBT is not rejected.

Table 1. MLEs of the parameters, the corresponding $\mathcal{L}$, AIC, K-S and p-value.

Model	$(\hat{\alpha}, \hat{\beta})$	$\mathcal{L}$	AIC	K-S	p-value
TPBT	$(4.6334 \times 10^{-5}, 0.4009)$	-392.0392	788.0784	0.1420	0.1373
Weibull	$(1.7946 \times 10^{-8}, 3.1133)$	-401.9936	807.9873	0.1821	0.0250

To show that there is a unique maximum value for the likelihood function of β, for every model, the plot of the profile log-likelihood function of β, $\mathcal{L}_\beta(\beta)$, for each model is provided in Figure 3. Also, the plot of log-relative

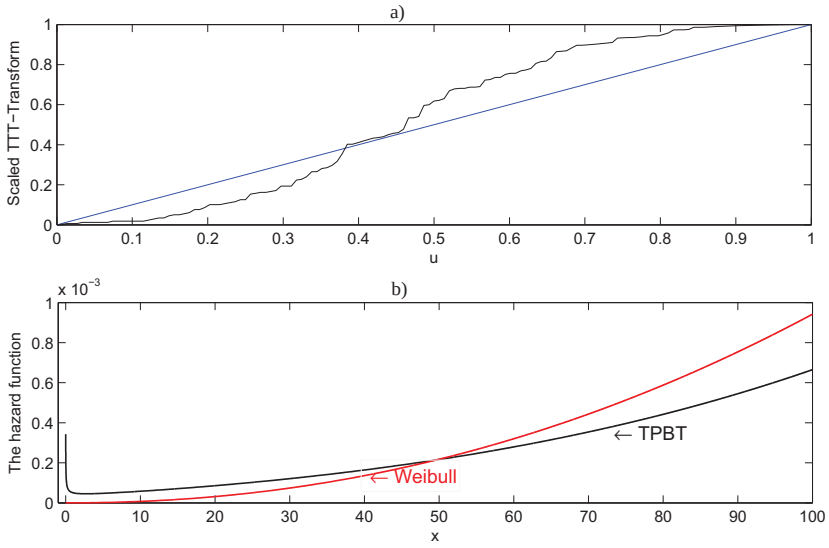


Figure 1. a) The Scaled TTT-Transform and b) the parametric hazard functions.

likelihood function of β, $r(\beta)$, for each model which can be used to decide on the plausible values of β, is provided in Figure 3.

Table 2 shows the 14.7% LI and 95% asymptotic CI for the TPBT and Weibull model parameters. The LI and asymptotic CI are not very similar simply because the relative log-likelihood functions of the parameters are not symmetric. The widths of the LI and Asymptotic interval for α in the TPBT model are 17.26×10^{-5} and 2.37×10^{-5} and for β are 0.0504 and 0.0087 respectively. These indicate that the asymptotic interval is more precise than the LI for both parameters.

Table 2. Confidence intervals.

Parameter	LI	Asymptotic
	TPBT	
α	$\left[9.9023\times10^{-6}, 1.8252\times10^{-4}\right]$	$\left[3.6282\times10^{-5}, 5.995\times10^{-5}\right]$
β	$[0.3752, 0.4256]$	$[0.3964, 0.4051]$
	Weibull	
α	$[3.7929\times10^{-10}, 4.5676\times10^{-7}]$	$[1.4057\times10^{-8}, 2.291\times10^{-8}]$
β	$[2.5465, 3.7855]$	$[3.0707, 3.1560]$

We used the three Bayesian techniques RS, MCMC and SB to estimate the TPBT parameters. The point estimates and the Bayesian intervals with the corresponding widths are shown in Table 3. The acceptance rates for SB, MCMC and RS are 21.288%, 64.905% and 50.66%, respectively. Figure 4 displays the simulated draws from the joint posterior distribution along with the contours of the likelihood function from the MCMC and SB procedures and the simulated draws using the RS algorithm on the contour plot of the posterior density. These plots show: (1) a good agreement between the likelihood function and the simulated draws from the MCMC; (2) the simulated draws from the SB are not well distributed within the likelihood contours; (3) the draws from the posterior distribution using the RS are well distributed on the contours of the exact posterior density. To compare the performance of the MCMC and SB to RS, we plotted the marginal posterior density functions using the three methods in Figure 5. From Figure 5, one can see that: (1) MCMC gives better estimate of the posterior densities than the SB; (2) SB provides an approximation to the posterior density with very small spread which explains why the interval estimates using SB are narrower than those obtained from RS and MCMC; (3) the posterior distribution of α is right skewed, while that of β is approximately symmetric.

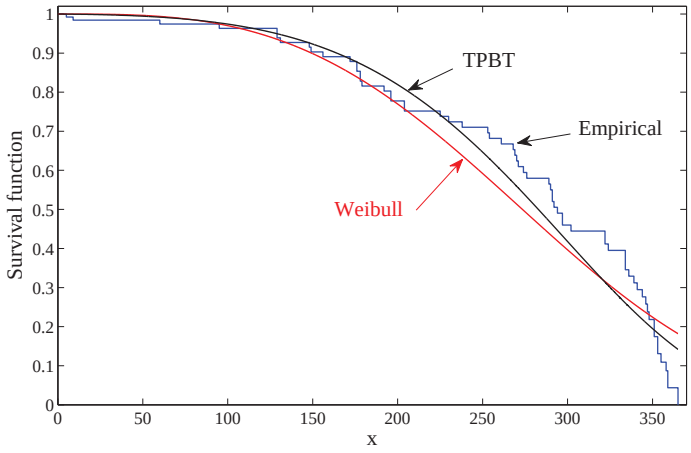


Figure 2. The empirical and parametric survival functions of the serum.

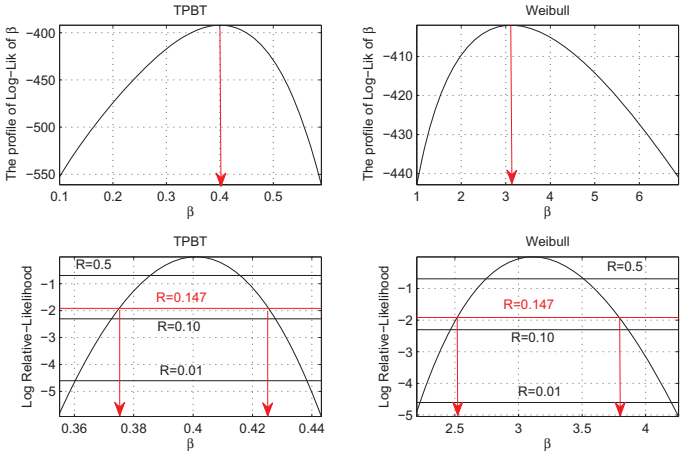


Figure 3. The profile log-likelihood function of β and the log-relative likelihood of β for both TPBT and Weibull models.

Table 3. Bayesian results for the TPBT model.

Method	Parameter	Point estimate	Bayesian interval	Width
SB	α	4.624×10^{-5}	$(3.559\times10^{-5}, 5.803\times10^{-5})$	2.245×10^{-5}
	β	0.40094	(0.4001, 0.4017)	0.0070
MCMC	α	5.333×10^{-5}	$(1.711\times10^{-5}, 1.253\times10^{-4})$	1.082×10^{-4}
	β	0.4006	(0.3835, 0.4177)	0.0341
RS	α	4.717×10^{-5}	$(1.331\times10^{-5}, 1.598\times10^{-4})$	1.4648×10^{-4}
	β	0.3998	(0.3780, 0.4207)	0.0427

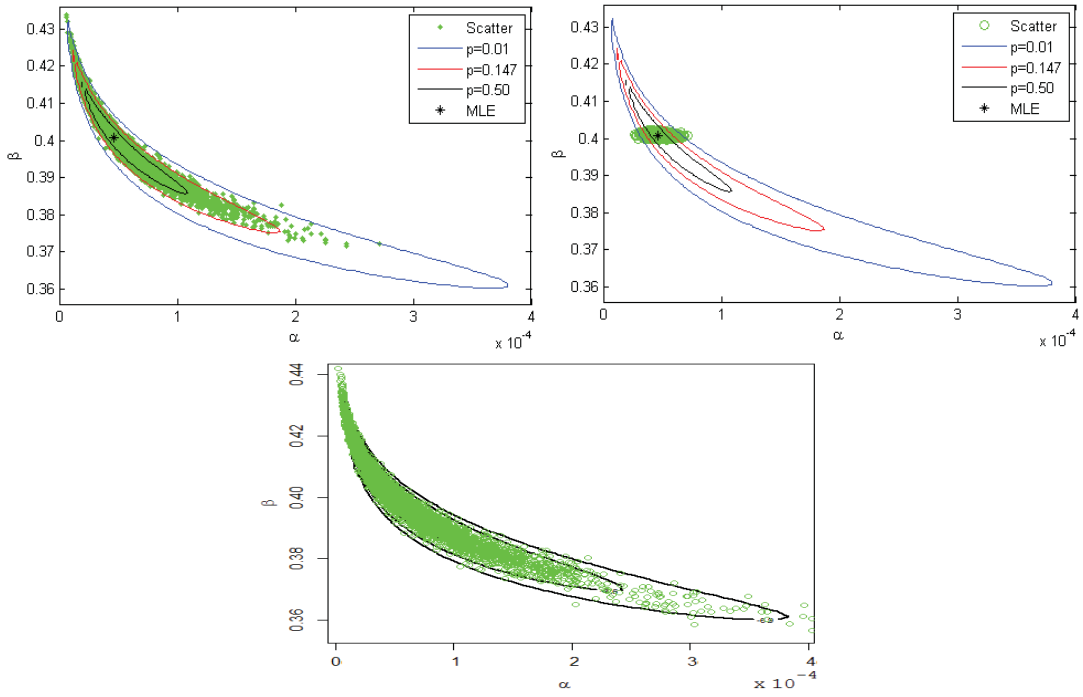


Figure 4. The 100p% likelihood contours along with the simulated draws of (α,β) from MCMC (top left) and SB (top right) and the simulated draws from the posterior density on contour plot of (α,β) using RS (bottom) for serum data.

7.2 Application 2: Aarset Data

Aarset data set consists of the failure times of 50 devices put on a life test. This data set was originally analyzed by Aarset (1987). Ng (2005) simulated progressively Type-II censored data based on the Aarset data as $\mathbf{x}$ = (0.1, 0.2, 1, 1, 1, 1, 1, 2, 3, 6, 7, 11, 18, 18, 18, 18, 21, 32, 36, 45, 47, 50, 55, 60, 63, 63, 67, 67, 75, 79, 82, 84, 84, 85, 86) and $\mathbf{r}$ = (0, 0, 0, 3, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0). Wu (2008) assumed that the data have the TPBT(α,β) distribution and obtained the MLE of the two parameters α and β, the 95% CI for only β (as shown in Table **??**) and a 95% confidence region for (α,β). The 95% confidence region is (Wu; 2008)

$$0.2147 < \beta < 0.4772,\ \frac{46.2248}{2\sum_{i=1}^{m}(r_i+1)\left(e^{x_i^{\beta}}-1\right)} < \alpha < \frac{90.0941}{2\sum_{i=1}^{m}(r_i+1)\left(e^{x_i^{\beta}}-1\right)}$$

We obtained the MLE and the corresponding standard error (se), likelihood intervals, likelihood regions and asymptotic confidence intervals for the TPBT distribution parameters. The inverse of the information matrix using the complete data and progressively Type-II censored data, respectively, are

$$\begin{bmatrix} 6.51965\times10^{-5} & -1.5925\times10^{-4} \\ -1.5925\times10^{-4} & 4.39656\times10^{-4} \end{bmatrix} \text{ and } \begin{bmatrix} 1.67091\times10^{-4} & -2.93646\times10^{-4} \\ -2.93646\times10^{-4} & 6.21791\times10^{-4} \end{bmatrix}$$

Also, we computed Bayes estimates and the two sided Bayesian intervals and their widths, using SB, MCMC and RS techniques, for the complete and progressively Type-II Aarset data as shown in Table 4. Figure 6 displays the simulated draws from the joint posterior distribution, using the MCMC procedure, on the contours of the likelihood function and the 14.7% likelihood region of (α,β) and the simulated draws from the posterior density, using RS, on the contour plot of (α,β), based on the complete and progressively Type-II censored Aarset data. We did not include the plot of the simulated draws obtained from SB in Figure 6 because, as we noticed in the first example, it does not provide representative samples.

Based on the results in Table 4 and plots in Figure 6, we can conclude that: (1) Bayesian intervals are more precise than the asymptotic and likelihood intervals; (2) MCMC gives more reasonable approximations to the

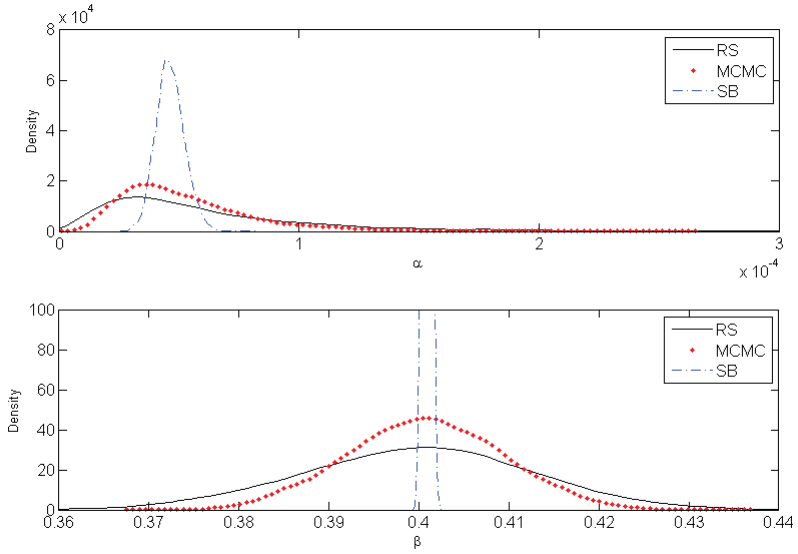


Figure 5. The marginal posterior density function using the three methods for serum data.

Bayes estimates than the SB; and (3) As expected, the complete version of the data gives more precise results than the progressively Type-II censored version.

Table 4. MLE and Bayes results for the Aarset data.

			Complete	Progressively censored
MLE	Point estimate(se)	α:	0.019381(0.008074435)	0.0315359(0.01292637)
		β:	0.347272(0.02096798)	0.310619 (0.02493574)
	95% CI	α:	[0.0035553, 0.0352065]	[0.00620065, 0.0568711]
	(width)		0.0316512	0.05067045
		β:	[0.3061760, 0.3883690]	[0.261746, 0.359492]
	(width)		0.082193	0.097746
	14.7% LI	α:	[0.0085446, 0.0397264]	[0.0142684, 0.061713]
	(width)		0.0311818	0.0474446
		β:	[0.3049220, 0.3870650]	[0.260251, 0.357761]
	(width)		0.082143	0.09751
Bayes				
SB	Point estimate	α:	0.0194307	0.0316467
		β:	0.347281	0.310603
	95%TBPI	α:	[0.01548, 0.0238742]	[0.0246732, 0.0395873]
	(width)		0.0083942	0.0149141
		β:	[0.344872, 0.349751]	[0.306733, 0.31445]
	(width)		0.004879	0.007717
MCMC	Point estimate	α:	0.0201862	0.0327917
		β:	0.346937	0.310308
	95%TBPI	α:	[0.0110438, 0.0336883]	[0.0177349, 0.0540883]
	(width)		0.0226445	0.0363534
		β:	[0.319192, 0.375192]	[0.277445, 0.344355]
	(width)		0.056	0.06691
RS	Point estimate	α:	0.01956508	0.0348011
		β:	0.34516	0.3072039
	95%TBPI	α:	[0.009702193, 0.03812835]	[0.01600871, 0.06126564]
	(width)		0.02842616	0.04525693
		β:	[0.3103567, 0.3787636]	[0.2620167, 0.3467852]
	(width)		0.0684069	0.08476856

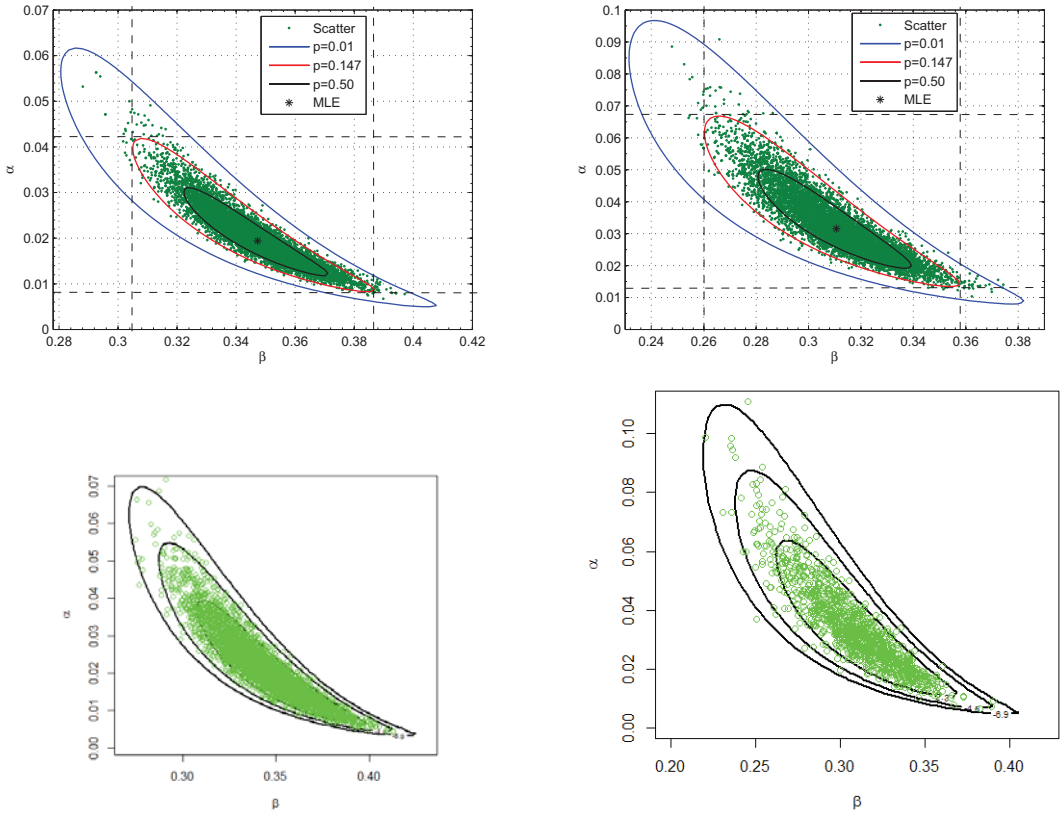


Figure 6. The 100p% likelihood contours along with the simulated draws of (α,β) from MCMC (top row) and the simulated draws from the posterior distribution of (α,β) on the contour plot using RS (bottom row) for complete (left column) and progressively censored (right column) Aarset data.

7.3 Simulated Data

To compare our proposed technique with technique in Wu (2008), we use the same simulated data in Wu (2008). Wu (2008) generated a progressively Type-II censored sample with $n = 15$ from the TPBT distribution with $\alpha = 0.02$ and $\beta = 0.5$, with censoring scheme $\mathbf{r}$ = (1,0,2,0,0,0,1,0,1,0) and got the observations $\mathbf{x}$ = (0.78, 3.15, 5.15, 6.69, 7.09, 7.40, 14.28, 15.72, 15.92, 22.59). Wu (2008) calculated the 95% CI for β as given in table **??**, and the 95% confidence region as

$$0.2299 < \beta < 0.7075, \quad \frac{8.5737}{2\sum_{i=1}^{10}(r_i+1)\left(e^{x_i^{\beta}}-1\right)} < \alpha < \frac{36.7141}{2\sum_{i=1}^{10}(r_i+1)\left(e^{x_i^{\beta}}-1\right)}$$

Table 5 shows the results obtained from the maximum likelihood and Bayes techniques. Based on the percentage errors, MCMC gives better estimation than both SB and RS algorithms for this simulated data set. The inverse of the information matrix is

$$\hat{\jmath}^{-1} = \begin{bmatrix} 3.2484\times10^{-4} & -1.00961\times10^{-3} \\ 1.00961\times10^{-3} & 3.8504\times10^{-3} \end{bmatrix}$$

Figure 7 displays The likelihood contours, the 95% confidence region along with the simulated draws from the posterior distribution using MCMC and SB, contour plot of the parameters (α,β) of the posterior density and of the transformed parameters $(\ln\alpha,\beta)$. The contour plot of (α,β) shows the skewness in the posterior density, especially towards when α gets larger. This is why we consider the transformation $(\ln\alpha,\beta)$. The skewness has been reduced and the distribution becomes more symmetric as shown in the bottom right plot in Figure 7. We also plot the simulated draws from the RS on the contour plot of the log posterior density in Figure 7. As expected, most of the draws fall within the contour of the exact posterior density.

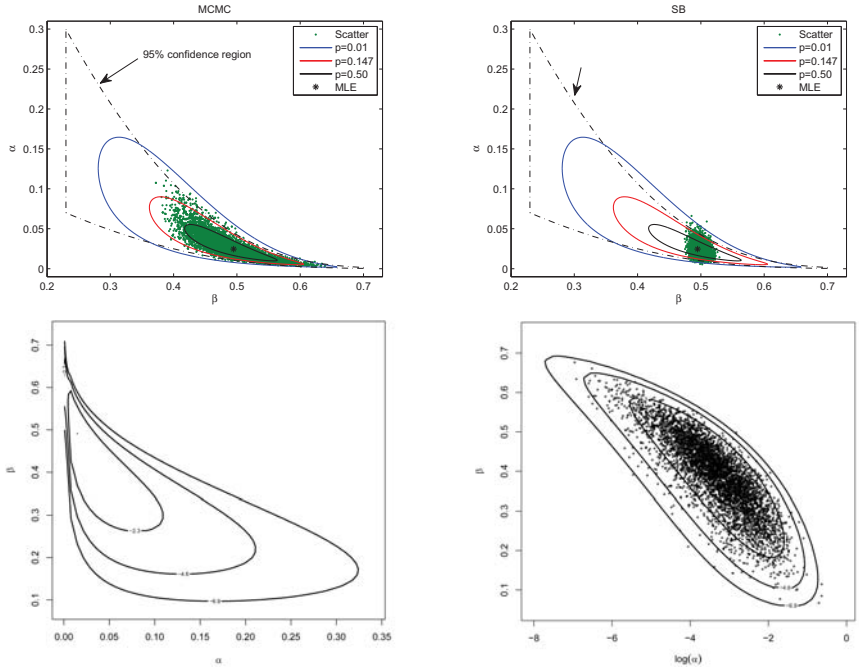


Figure 7. The 100p% likelihood contours, 95% confidence region along with the simulated draws of (α,β) using MCMC (top left) and SB (top right), contour plot of (α,β) (bottom left) and of $(\log\alpha,\beta)$ (bottom right) along with simulated draws from the posterior density using RS for the progressively censored simulated data.

Table 5. Point estimates and the corresponding percentage errors and interval estimates of the parameters for the simulated data.

Method	Statistic	α	β
MLE	Point Estimate	0.0245174	0.494714
	Percentage Error	22.5869	1.05723
	Wu's 95% CI	–	[0.2551, 0.6785]
	width	–	0.423366
	Approximate 95% CI	(0, 0.0598424]	[0.373095, 0.616333]
		0.0598424	0.243238
	14.7% LI	[0.00564596, 0.076716]	[0.361701, 0.605369]
	width	0.07107004	0.243668
Bayes			
SB	Point Estimate	0.0243745	0.494892
	Percentage Error	21.8725	1.0216
	95% TBPI	[0.0114819, 0.0387302]	[0.48498, 0.514757]
	width	0.0272483	0.029777
MCMC	Point Estimate	0.0229747	0.500086
	Percentage Error	14.8737	0.0172637
	95% TBPI	[0.00728443, 0.0627655]	[0.417348, 0.57805]
	width	0.05548107	0.160702
RS	Point Estimate	0.01653803	0.4907907
	Percentage Error	17.3099	1.84186
	95% TBPI	[0.003723937, 0.0628734]	[0.3717111, 0.6053621]
	width	0.05914946	0.233651

8. Simulation Studies

To evaluate the performance of maximum likelihood and Bayes procedures (using the SB, MCMC and RS techniques) based on the sample and the censored sizes, a large simulation study using Monte Carlo method is carried

out according to the following scheme:

1. Specify the sample size n.
2. Specify the censored size $n - m$ as a percentage of n.
3. Generate a random sample with sizes n and m from TPBT(α, β).
4. Compute MLEs of α and β.
5. Compute the Fisher information matrix at the MLE's of α and β.
6. Compute Bayes estimates of α and β using SB, MCMC, and RS methods.
7. Compute the squared deviation of the point estimate of each parameter using each of the four procedures from the corresponding true value.
8. Repeat Steps 3-7 1000 times.
9. Calculate the average of the point estimates of every parameter and the mean squared error (MSE) associated with each estimate for the two parameters using the four techniques.
10. Steps 2-9 are performed when $n - m = 0\%, 10\%, 20\%, \cdots, 70\%$ of n.
11. Steps 1-10 are performed when $n = 10, 15, 20, \cdots, 55$.
12. Steps 1-11 are carried out when $\alpha = 0.1$ and $\beta = 0.5$.

Table 6 presents the average estimates (first row) and the corresponding MSE (second row) corresponding to every parameter at different set of values of (n, m).

Based on the simulation results, one can conclude that: (1) the MSE decreases when n increases for all cases at every level of censoring, (2) the MSE increases when the percentage of censorship increases mainly for large n, (3) there is no significant difference in the values of the MSE corresponding to the three techniques. However, based on the MSE, (*i*) the MCMC generally provides better approximations than SB; (*ii*) RS produces 75.62% better estimations than MCMC.

9. Conclusion

This paper extends the work of Sarhan et al. (2012) which considered a two parameter bathtub shaped distribution that was revisited by Chen (2000). We provide statistical inference of the two parameters using maximum likelihood and Bayesian methods. For Bayesian, the simulation based, Markov Chain Monte Carlo and rejection sampling techniques are applied. The specific interest is in the MCMC and RS methods. The estimation techniques were applied to two real data sets and a simulated data set. In addition, the estimation methods were compared by a Monte Carlo simulation.

For the two real data sets, the Bayesian estimates of the parameters from the MCMC and RS methods were not very similar to those obtained by the SB method. The widths of the Bayesian intervals for the parameters constructed by the MCMC and RS methods were larger. In each parameter case, the widths of the MCMC and RS intervals are about 5 times larger than the SB interval widths. The main reason was that the updates from the SB algorithm are not representative.

Several phenomena can be observed from the results from the Monte Carlo simulation. First the mean squared error decreases as the sample size increases for fixed level of the number of units censored. For a fixed value of the sample size beyond some threshold value, the mean squared error increases with increasing level of units censored. The RS method gives better approximations of the point estimates of the parameters as the corresponding mean squared errors are generally smaller than those from the other methods.

Table 6. The mean estimated values and corresponding MSE using MLE and Bayes technique methods.

n	m	MLE		Bayes SB		Bayes MCMC		Bayes RS	
		α	β	α	β	α	β	α	β
10	10	0.0965938	0.543229	0.0977269	0.54325	0.104814	0.539614	0.108399602	0.537720715
		0.00352657	0.0101639	0.00359058	0.0101641	0.00378798	0.00975912	0.003634698	0.010989128
	9	0.088357	0.544541	0.0894954	0.544532	0.0966724	0.540575	0.101505992	0.526144192
		0.00330554	0.0116283	0.00336196	0.011625	0.00342947	0.0112017	0.003590120	0.009536702
	8	0.0796068	0.54592	0.0807074	0.54594	0.0878191	0.541141	0.096275747	0.515060776
		0.00343309	0.0132877	0.00346612	0.0132822	0.00339047	0.0127352	0.003175031	0.007210631
	7	0.0688228	0.547092	0.0699216	0.547092	0.0768808	0.542017	0.086356682	0.516791911
		0.00313645	0.0139212	0.00313905	0.0139168	0.00290339	0.0133409	0.002287333	0.013380657
	6	0.0620851	0.54514	0.0632252	0.545137	0.07002	0.539505	0.070573258	0.516746353
		0.00355225	0.0174726	0.00354032	0.017462	0.0032198	0.0167929	0.001879280	0.007272028
	5	0.0538699	0.544946	0.0550502	0.544956	0.0615898	0.538361	0.055949112	0.550375269
		0.00384942	0.0219228	0.00381704	0.0219072	0.00337681	0.0210178	0.003134624	0.017955789
	4	0.0438866	0.546227	0.0450433	0.546268	0.0513286	0.53851	0.048060585	0.535066812
		0.00449851	0.0254716	0.00444282	0.0254549	0.00388854	0.0245151	0.003836292	0.020156080
	3	0.032328	0.55412	0.0335084	0.554039	0.039321	0.544672	0.032985715	0.546116807
		0.00536857	0.0309052	0.00527079	0.0309288	0.00460875	0.0295594	0.004838608	0.015649686
15	15	0.100184	0.525084	0.100935	0.525087	0.10605	0.522522	0.103980542	0.529920484
		0.00225603	0.00528135	0.00228534	0.0052806	0.00240274	0.00511315	0.002617119	0.007763336
	14	0.0909943	0.530283	0.0917494	0.530284	0.0968113	0.527427	0.104149751	0.516204702
		0.00209798	0.00640082	0.00211125	0.00640029	0.00213668	0.00618487	0.002845662	0.006203468
	12	0.0783714	0.531331	0.0791134	0.531331	0.0840893	0.528149	0.089883498	0.513820595
		0.00226732	0.00719851	0.00226286	0.00719396	0.00216391	0.00694955	0.001684704	0.005211331
	11	0.0721472	0.526631	0.0729103	0.526627	0.077784	0.523216	0.079724604	0.502702003
		0.00225351	0.00698954	0.00224076	0.00698635	0.00206838	0.00676461	0.001860149	0.008583772
	9	0.0615553	0.529293	0.0623289	0.529297	0.067142	0.525219	0.066779047	0.516578941
		0.00283782	0.00997	0.00281112	0.00996494	0.0025335	0.00964735	0.002150986	0.006713115
	8	0.0554091	0.528275	0.0561979	0.528268	0.0609722	0.523734	0.061997086	0.510289929
		0.00310942	0.0113079	0.003071	0.0112997	0.00273604	0.0109581	0.002573363	0.008069084
	6	0.0405175	0.533492	0.0412645	0.533493	0.0457878	0.527877	0.048064553	0.508517583
		0.00429142	0.0135554	0.00423043	0.0135447	0.00377333	0.0130829	0.003283371	0.008654256
	5	0.0367048	0.530218	0.0375033	0.530232	0.0419314	0.523801	0.030302738	0.535244630
		0.00475745	0.0183476	0.00468999	0.0183348	0.00420218	0.0177115	0.005154421	0.014734452
20	20	0.100025	0.520083	0.100546	0.520087	0.104492	0.518093	0.110610084	0.506698229
		0.002008	0.00397173	0.00202168	0.00397055	0.002108	0.00388291	0.002264086	0.003449096
	18	0.0901603	0.517461	0.0907292	0.517449	0.094601	0.515299	0.097528212	0.510816129
		0.00160633	0.00389142	0.00160836	0.00389002	0.00160002	0.00379748	0.001500577	0.004353692
	16	0.0780805	0.521752	0.0786473	0.521749	0.0823812	0.519374	0.089504714	0.505521191
		0.00181521	0.0048663	0.00180555	0.00486482	0.00170788	0.00474747	0.001520748	0.004244547
	14	0.0688834	0.523839	0.0694442	0.523841	0.0731434	0.521107	0.075618078	0.511536977
		0.00213282	0.00606386	0.00211471	0.00606354	0.00194819	0.00591089	0.001764872	0.005174954
	12	0.0585814	0.522823	0.0591269	0.52282	0.0628395	0.519677	0.068418164	0.507210644
		0.00260232	0.00624033	0.00257161	0.00623965	0.00232439	0.00607496	0.001996491	0.005806728
	10	0.0516951	0.518968	0.0522706	0.518957	0.0559389	0.515271	0.058972871	0.504314342
		0.00318661	0.00813416	0.00314724	0.00813041	0.00284923	0.00792296	0.002485735	0.007082680
	8	0.0416027	0.520105	0.0422094	0.520129	0.0457422	0.515628	0.049264789	0.502262124
		0.00398774	0.00939556	0.00393447	0.00939277	0.0035661	0.00914802	0.003227021	0.009296318
	6	0.0311785	0.524733	0.0317564	0.524733	0.0351537	0.519104	0.038595322	0.496787510
	6	0.00515314	0.0132197	0.00508979	0.0132057	0.00466518	0.0128178	0.004166036	0.010517184
25	25	0.0979555	0.517154	0.0983677	0.517147	0.101477	0.515601	0.103888994	0.510443337
		0.00143602	0.00309556	0.00144151	0.00309435	0.00147783	0.00302777	0.001439089	0.002989788
	23	0.0911666	0.516757	0.0915688	0.516752	0.0947404	0.515049	0.093075773	0.511015683
		0.00142217	0.00355025	0.00142239	0.00354986	0.00141908	0.00348646	0.001348822	0.003458540
	20	0.0779644	0.518965	0.0783778	0.518969	0.0814824	0.516957	0.086790467	0.506201220
		0.00154373	0.0037997	0.00153332	0.0037991	0.00144077	0.00370779	0.001188499	0.003196711
	18	0.0688486	0.523022	0.0692769	0.523026	0.0722687	0.520835	0.079904018	0.504490909
		0.00185523	0.00435935	0.00183745	0.00435874	0.0016931	0.00423545	0.001366548	0.003535451
	15	0.0599343	0.517435	0.0603883	0.517439	0.063386	0.514951	0.067194003	0.504849203
		0.00236278	0.00506006	0.00233574	0.0050565	0.00213315	0.00494708	0.001843134	0.004277783
	13	0.0519465	0.51907	0.0524132	0.519074	0.0553559	0.516249	0.056684709	0.497956849
		0.00299153	0.00608319	0.0029569	0.00608104	0.00271389	0.00595449	0.002489124	0.005447762
	10	0.0417998	0.516557	0.0422673	0.516556	0.0451913	0.512843	0.047737408	0.496585558
		0.0038927	0.00714924	0.00384858	0.00714308	0.00354092	0.00697894	0.003186006	0.006461775
	8	0.0333519	0.518054	0.0338114	0.518048	0.0366399	0.513668	0.039214367	0.495359818
		0.00483082	0.0095539	0.00477984	0.00955074	0.00443303	0.00931808	0.004065767	0.007788294
30	30	0.0972154	0.515079	0.0975349	0.515085	0.10021	0.513688	0.1060592628	0.506283227
		0.00111359	0.00245533	0.00111618	0.00245524	0.00113271	0.00240554	0.0012332344	0.002325400
	27	0.0887368	0.51338	0.0890602	0.51338	0.0916946	0.511908	0.0953604072	0.505068810
		0.00111084	0.00256706	0.0011075	0.00256683	0.00108281	0.00253086	0.0011058401	0.002363809
	24	0.0807674	0.511062	0.0811108	0.511066	0.0838238	0.509323	0.0850070459	0.505923054
		0.00121665	0.00264025	0.00120795	0.00263942	0.00113795	0.00259585	0.0010669811	0.002755070
	21	0.0704814	0.51272	0.070829	0.512721	0.0734354	0.510873	0.0763238677	0.502793470
		0.00163933	0.0035076	0.00162409	0.00350678	0.00150204	0.00345702	0.0013449374	0.003226490
	18	0.0602792	0.513873	0.0606465	0.51387	0.0632074	0.511759	0.0665972335	0.501458937
		0.00220679	0.00399115	0.0021832	0.00398996	0.0020086	0.00391676	0.0017751811	0.003921401
	15	0.0505121	0.514852	0.0508903	0.514848	0.0533945	0.512296	0.0570783302	0.499624204
		0.00296213	0.00494479	0.00293033	0.00494297	0.00270905	0.00483695	0.0024065128	0.004355206
	12	0.0411194	0.511906	0.0415022	0.511914	0.044003	0.508708	0.0473113451	0.495747118
		0.0038711	0.0059779	0.00383284	0.00597479	0.00356323	0.00586353	0.0031819949	0.005362933
	9	0.0319218	0.510458	0.0323116	0.510445	0.0347173	0.506466	0.0357581412	0.499010244
		0.00495134	0.00796539	0.0049055	0.0079599	0.00460138	0.00782591	0.0044219334	0.007251997
35	35	0.10033	0.510093	0.100576	0.510096	0.102921	0.508967	0.1050598843	0.504776950
		0.00102623	0.00204779	0.00102869	0.00204781	0.00105844	0.00202386	0.0009980059	0.001837138
	32	0.0920323	0.510244	0.0922977	0.510249	0.0946753	0.508911	0.0957556311	0.505687660
		0.00101627	0.00226071	0.00101504	0.00226079	0.00100571	0.00223067	0.0009993649	0.002263441
	28	0.0805466	0.509987	0.0808194	0.509989	0.0831057	0.508521	0.0860528455	0.501269403
		0.00113601	0.00245633	0.0011274	0.00245604	0.00106022	0.00241068	0.0009560736	0.002318271
	25	0.0710553	0.511721	0.071354	0.511728	0.0736569	0.510039	0.0765145830	0.498257832
		0.00147962	0.00273333	0.00146566	0.00273284	0.00135798	0.00268661	0.0012302975	0.002742438
	21	0.0604706	0.510207	0.0607729	0.510205	0.0630212	0.508302	0.0650159865	0.501368529
		0.00208918	0.00321931	0.00206922	0.00321875	0.00191287	0.00317459	0.0017600030	0.002939498
	18	0.0529151	0.509529	0.0532462	0.509529	0.0554463	0.507349	0.0582921469	0.496170424
	18	0.0027137	0.00414855	0.0026869	0.00414556	0.00250154	0.00409144	0.0022509046	0.003775513
	14	0.0410745	0.510033	0.0414086	0.510038	0.0435476	0.507386	0.0462818421	0.496623113
		0.00380462	0.0046829	0.0037698	0.00468083	0.00353613	0.0046115	0.0032510249	0.004686905
	11	0.0330037	0.5108	0.0333461	0.510807	0.0354775	0.507396	0.0347878456	0.490382611
		0.00476481	0.00637757	0.00472357	0.00637076	0.00445849	0.0062901	0.0045029202	0.006254630
40	40	0.099592	0.50994	0.0998086	0.509943	0.101859	0.508928	0.1027110237	0.505816211
		0.000915641	0.00173877	0.000916745	0.00173886	0.000938836	0.00171816	0.0008710900	0.001739452
	36	0.0896171	0.511357	0.0898316	0.511361	0.0919229	0.510193	0.0936502366	0.504116166
		0.00092654	0.00210103	0.000924203	0.00210072	0.000903531	0.00206905	0.0007803555	0.001802891
	32	0.0796949	0.510414	0.0799226	0.510421	0.0819627	0.509117	0.0843232337	0.503439167
		0.00108337	0.00226917	0.00107639	0.00226871	0.00101375	0.00224361	0.0009535063	0.001986631
	28	0.070364	0.510544	0.0706078	0.510542	0.0726423	0.509069	0.0743081046	0.502956681
		0.00145944	0.00241471	0.00144698	0.00241445	0.00134653	0.00238228	0.0012230970	0.002212039

n	m	MLE		Bayes SB		Bayes MCMC		Bayes RS	
		α	β	α	β	α	β	α	β
	24	0.0589237	0.511938	0.0591778	0.511944	0.0611501	0.510222	0.0637834036	0.503850526
		0.00212529	0.0028013	0.00210674	0.00280073	0.00196249	0.0027554	0.0017856161	0.002760099
	20	0.0508943	0.508294	0.0511654	0.508289	0.0531312	0.506295	0.0545042108	0.500074972
		0.00276204	0.00325971	0.00273818	0.00325978	0.00255928	0.00322043	0.0024751614	0.003288732
	16	0.0407576	0.51077	0.0410403	0.510777	0.0429512	0.508298	0.0448721380	0.498912840
		0.00381568	0.00453387	0.0037857	0.00453127	0.00357632	0.00446909	0.0033527216	0.004070938
	12	0.0313449	0.507069	0.0316441	0.507074	0.0335126	0.503964	0.0350569184	0.494194384
		0.00493557	0.00519936	0.00489815	0.00519722	0.0046553	0.00512012	0.0044477989	0.005417482
45	45	0.099404	0.509762	0.0995777	0.509761	0.101452	0.508819	0.1034950866	0.504796789
		0.000700984	0.00147888	0.000701514	0.00147869	0.000715174	0.00145551	0.0007970579	0.001383290
	41	0.0905794	0.508827	0.0907641	0.50883	0.0925778	0.507828	0.0919737663	0.503814669
		0.000774688	0.00169698	0.000772756	0.00169676	0.000753132	0.0016712	0.0007955076	0.001582602
	36	0.0804721	0.507271	0.0806633	0.507273	0.0824927	0.506152	0.0847056266	0.501305058
		0.00100127	0.00193854	0.00099529	0.00193842	0.000939194	0.00191766	0.0008736683	0.001776347
	32	0.071167	0.507539	0.0713656	0.507544	0.0732112	0.506228	0.0011178763	0.002072463
		0.00131079	0.00204462	0.0013009	0.00204486	0.00121075	0.00202527	0.0758102308	0.502013931
	27	0.0596875	0.510219	0.0599065	0.510229	0.0616798	0.50867	0.0638147000	0.504241774
		0.00202708	0.00241344	0.00201171	0.00241312	0.00188221	0.00237689	0.0017339790	0.002463779
	23	0.0525129	0.506082	0.0527449	0.506091	0.0545259	0.504284	0.0532987239	0.498539769
		0.0026475	0.00300329	0.00262761	0.00300276	0.00247301	0.00297257	0.0025246835	0.002869696
	18	0.0396012	0.512335	0.0398377	0.512338	0.0415572	0.510104	0.0453582536	0.493997303
		0.00391486	0.00394246	0.00388864	0.00394025	0.00369323	0.00386819	0.0032401918	0.003322612
	14	0.0311549	0.511945	0.0314102	0.511951	0.0330341	0.509314	0.0361276242	0.493710912
		0.00492781	0.00473281	0.0048952	0.00472983	0.00468179	0.00464864	0.0042821305	0.004457198
50	50	0.0995724	0.508182	0.0997171	0.508187	0.101381	0.50738	0.1035335180	0.503039498
		0.000655787	0.00136083	0.000656443	0.00136084	0.000665895	0.00134662	0.0007241307	0.001301183
	45	0.0904489	0.507345	0.090604	0.507347	0.092247	0.506425	0.0943669027	0.501332417
		0.000714751	0.00148269	0.000712485	0.00148247	0.00069268	0.00146682	0.0006399489	0.001346615
	40	0.0781412	0.510993	0.0782879	0.510994	0.0799672	0.509953	0.0832603379	0.504195263
		0.000971694	0.00179197	0.000966141	0.00179182	0.000907772	0.0017728	0.0007927074	0.001652673
	35	0.069269	0.509038	0.0694478	0.509039	0.0710814	0.507872	0.0751999064	0.499655530
		0.00135772	0.00186597	0.001348	0.00186576	0.00126008	0.0018419	0.0011150765	0.001811460
	30	0.0591685	0.509075	0.0593578	0.509078	0.060997	0.507634	0.0655125831	0.497755539
		0.00201064	0.00215451	0.00199666	0.00215454	0.00187466	0.00212469	0.0015827993	0.001960856
	25	0.0503399	0.507663	0.0505373	0.507669	0.0521508	0.506057	0.0547102650	0.499009974
		0.00275415	0.00260065	0.00273601	0.00259974	0.00258774	0.00256945	0.0023687057	0.002506362
	20	0.0399768	0.511354	0.040192	0.511351	0.0417161	0.509403	0.0439064657	0.498555889
		0.00384476	0.00348622	0.00382065	0.00348563	0.00364728	0.00343041	0.0033724790	0.003115709
	15	0.0308155	0.506553	0.0310507	0.506561	0.032579	0.503916	0.0343540871	0.496051442
		0.00496856	0.00451572	0.0049381	0.00451354	0.00473602	0.00445921	0.0044776877	0.004107279
55	55	0.0999156	0.507076	0.100043	0.50708	0.101586	0.506317	0.1025671133	0.504709426
		0.000605656	0.00118931	0.000606407	0.00118933	0.000618069	0.0011795	0.0006159297	0.001119474
	50	0.0896197	0.508431	0.0897495	0.508432	0.0912528	0.507659	0.0941894391	0.502748280
		0.000657108	0.00131065	0.000655005	0.00131069	0.000636228	0.00130038	0.0005941511	0.001231337
	44	0.0790223	0.508002	0.0791718	0.508003	0.0806547	0.507122	0.0830742540	0.502139363
		0.000896261	0.00149798	0.00089051	0.00149778	0.000839479	0.00148533	0.0007791093	0.001454397
	39	0.0713882	0.507398	0.0715437	0.507401	0.0730488	0.506403	0.0735465362	0.500035985
		0.00119633	0.00167253	0.00118811	0.00167254	0.00111315	0.00165792	0.0011417480	0.001741524
	33	0.0596584	0.507667	0.0598219	0.507673	0.0612978	0.506414	0.0646420676	0.498748827
		0.00193851	0.00197268	0.00192602	0.00197214	0.0018161	0.00194866	0.0016278846	0.002142306
	28	0.0510951	0.507843	0.0512697	0.507851	0.0527337	0.506403	0.0543184546	0.500949336
		0.00265338	0.00216282	0.00263717	0.00216224	0.00250405	0.00213783	0.0023899234	0.002317586
	22	0.0405079	0.506676	0.0407017	0.506683	0.0421262	0.504859	0.0438771961	0.497455014
		0.00374537	0.00289383	0.00372366	0.00289281	0.00356261	0.00285924	0.0033688235	0.002975789
	17	0.0311658	0.510179	0.0313687	0.510182	0.0327539	0.507893	0.0333519673	0.493075804
		0.00489078	0.00388375	0.00486439	0.00388256	0.00468169	0.00382626	0.0045952979	0.004124284

Acknowledgments

The authors would like to thank the referee for the valuable suggestion to use the rejection sampling, which improved the paper.

References

Albert, J. (2009). Bayesian Computation with R, Springer. http://dx.doi.org/10.1007/978-0-387-92298-0

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D. (2014). Bayesian Data Analysis, Third Edition, Chapman & Hall/CRC.

Chen, M.H., Shao, Q.M., and Ibrahim, J.G. (2000), Monte Carlo Methods in Bayesian Computation. Spring Series in Statistics, Vol 32, Springr. http://dx.doi.org/10.1007/978-1-4612-1276-8

Hastings, W. (1970), Monte Carlo sampling methods using Markov chain and their applications, Biometrika 57, 97-109. http://dx.doi.org/10.1093/biomet/57.1.97

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953), Equations of state calculations by fast computing machine, Journal of Chemical Physics 21, 1087-1092. http://dx.doi.org/10.1063/1.1699114

Chib, S. and Greenberg, E. (1995), Understanding the Metropolis-Hastings algorithm. American Statistician, 49: 327 - 335.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996), Markov Chain Monte Carlo in Practice, Chapman and Hall, London.

Tierney, L. (1994), Markov chains for exploring posterior distributions (with discussion). Annals of Statistics, 22: 1701-1762. http://dx.doi.org/10.1214/aos/1176325750

Silva, A.N.F.(2004), Estudo evolutivo das crianas expostas ao HIV e notificadas pelo ncleo de vigilncia epidemiolgica do HCFMRP-USP, MSc thesis, University of So Paulo, Brazil.

Perdon, G.S.C. (2006), Modelos de riscos aplicados anlise de sobrevivncia, Doctoral thesis, Institute of Computer Science and Mathematics, University of So Paulo, Brasil (in Portuguese).

Smith, R.M. and Bain, L.J. (1975), An Exponential Power Life-Testing Distribution, Communications in Statistics, 4 , 469-481. http://dx.doi.org/10.1080/03610927508827263

Leemis, L. (1986), Relationships Among Common Univariate Distributions, The American Statistician , 40 , 143 - 146.

Gaver, D.P. and Acar, M. (1979), Analytical hazard representations for use in reliability, mortality, and simulation studies, Communication in Statistics, B(Sim. & Comp.), 8, 91-111.

Hjorth, U. (1980), A reliability distribution with increasing, decreasing, and bathtub-shaped failure rates, Technometrics, 22, 99-107. http://dx.doi.org/10.2307/1268388

Chen, Z. (2000),A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function, Statistics and Probability Letters, 49, 155-161. http://dx.doi.org/10.1016/S0167-7152(00)00044-4

Gurvich, M.R., Dibenedetto, A.T. and Rande, S.V. (1997), A new statistical distribution for characterizing the random strength of brittle materials, Journal of Material Science, 32, 2559-2564. http://dx.doi.org/10.1023/A:1018594215963

Haynatzki, G.R., Weron, K. and Haynatzka, V.R. (2000), A new statistical model of tumor latency time, Mathematical and Computer Modeling, 32, 251-256. http://dx.doi.org/10.1016/S0895-7177(00)00132-1

Kalbfleisch, J.G., (1985), Probability and Statistical Inference: Volume 2: Statistical Inference, Springer Texts in Statistics.

Martz, H.F. and Waller, R.A. (1982), Bayesian Reliability Analysis, Wiley, New York.

Meeker, W.O. and Escobar, L.A. (1998), Statistical Methods for Reliability Data, John Wiley & Sons, Inc., New York.

Mudholkar, G.S. and Srivastava, D. K. (1993), Exponentiated Weibull Family for Analyzing Bathtub Failure-Rate Data, IEEE Trans. Rel., 42(2), 299-302. http://dx.doi.org/10.1109/24.229504

Nadaraja, S. and Kotz, S. (2007), The two-parameter bathtub-shaped lifetime distribution, Quality Reliability Engineering International, 23, 279-280. http://dx.doi.org/10.1002/qre.795

Ng, H.K.T. (2005), Parameter estimation for a modified Weibull distribution, for progressively type-II censored samples, IEEE Trans. Reliab. 54, 374-380. http://dx.doi.org/10.1109/TR.2005.853036

Pham, H. and Lai, C.D. (2007), On recent generalizations of the Weibull distribution, IEEE Trans. Reliab., 56, 454-458. http://dx.doi.org/10.1109/TR.2007.903352

Sarhan, A.M. and Apaloo, J. (2013), Exponentiated modified Weibull extension distribution, Reliability Engineering & System Safety, 112, 137-144. http://dx.doi.org/10.1016/j.ress.2012.10.013

Sarhan, A.M., Hamilton, D.C. and Smith, B. (2012), Parameter estimation for a two-parameter bathtub-shaped lifetime distribution, Applied Mathematical Modelling, 36, 5380-5392. http://dx.doi.org/10.1016/j.apm.2011.12.054

Sarhan, A.M., Ahmad, A.A. and Alasbahi, I.A., (2013), Exponentiated generalized linear exponential distribution, Applied Mathematical Modelling, 37-5, 2838-2849.

Wu, S.-J. (2008), Estimation of the two-parameter bathtub-shaped lifetime distribution with progressive censoring, Journal of Applied Statistics, 35(10), 1139-1150. http://dx.doi.org/10.1080/02664760802264996

Wu, J.-W., Lu, H.-L., Chen, C.-H., and Wu, C.-H. (2004), Statistical inference about the shape parameter of the new two-parameter bathtub-shaped lifetime distribution, Quality Reliab. Eng. Int. 20, 607-616. http://dx.doi.org/10.1002/qre.572

International Journal of Statistics and Probability; Vol. 4, No. 4; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Statistical Evaluation of Face Recognition Techniques under Variable Environmental Constraints

Louis Asiedu[1], Atinuke O. Adebanji[2], Francis Oduro[3] & Felix O. Mettle[4]

[1]Department of Statistics, University of Ghana, Legon-Accra, Ghana

[2]Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

[3]Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[4]Department of Statistics, University of Ghana, Legon-Accra, Ghana

Correspondence: Louis Asiedu, Department of Statistics, University of Ghana, Legon-Accra, Ghana. Tel: 233-543-426-707. E-mail: lasiedu@ug.edu.gh

Received: August 1, 2015 Accepted: August 19, 2015 Online Published: October 9, 2015

doi:10.5539/ijsp.v4n4p93 URL: http://dx.doi.org/10.5539/ijsp.v4n4p93

Abstract

Experiments have shown that, even one to three day old babies are able to distinguish between known faces (Chiara, Viola, Macchi, Cassia, & Leo, 2006). So how hard could it be for a computer? It has been established that face recognition is a dedicated process in the brain (Marque´s, 2010). Thus the idea of imitating this skill inherent in human beings by machines can be very rewarding though the idea of developing an intelligent and self-learning system may require supply of sufficient information to the machine. This study proposes multivariate statistical evaluation of the recognition performance of Principal Component Analysis and Singular Value Decomposition (PCA/SVD) and a Whitened Principal Component Analysis and Singular Value Decomposition algorithms (Whitened PCA/SVD) under varying environmental constraints. The Repeated Measures Design, Paired Comparison test, Box's M test and Profile Analysis were used for performance evaluation of the algorithms on the merit of efficiency and consistency in recognizing face images with variable facial expressions. The study results showed that, PCA/SVD is consistent and computationally efficient when compared to Whitened PCA/SVD.

Keywords: Principal Component Analysis, Singular Value Decomposition, whitening, multivariate**,** repeated measures design, Paired Comparison, Box's M and profile analysis.

1. Introduction

Face recognition is an easy task for humans. Although the ability to infer the intelligence or character from facial appearance is suspect, the human ability to recognize faces is remarkable (Turk & Pentland, 1991). According to Rahman (2013), the intricacy of a face features originate from continuous changes in the facial features that take place over time. Regardless of these changes, we are able to recognize a person very easily.

In recent years, face recognition techniques have gained significant attention from researchers partly because face recognition is non-invasive with a sense of primary identification. One of the main driving factors for face recognition is the ever growing number of applications that an efficient and resilient recognition technique addresses; for example, security systems based on biometric data, criminal identification, missing children identification, passport/driver license, voter identification and user-friendly human-machine interfaces. An example of the later category is smart rooms, which use cameras and microphones arrays to detect the presence of humans, decide on their identity and then react according to the predefined set of preferences for each person.

Currently, all face recognition techniques work in either of the two ways. One is local face recognition system which uses facial features (nose, mouth, eyes) of a face. That is to consider the fiducial points in the face to associate the face with a person. The local-feature method computes the descriptor from parts of the face and gathers information into one descriptor. Some local-feature methods are, Local Feature Analysis (LFA), Garbor Features, Elastic Bunch Graph Matching (EBGM) and Local Binary Pattern Feature Agrawal *et al.,* (2014).

The second approach or global face recognition system uses the whole face to identify a person. The principle of whole face method is to construct a subspace using Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Random Projection (RP), or Non-negative Matrix

Factorization (NMF). These are all dimensionality reduction algorithms that seek to reduce the large dimensional face image data to small dimension for matching.

Viola and Jones (2001) proposed a multi-stage classification procedure for face recognition that reduces the processing time substantially while achieving almost the same accuracy as compared to a much slower and more complex single stage classifier. Lienhart and Maydt (2002) extends their rapid object detection framework in two important ways: Firstly, their basic and over-complete set of haar-like feature was extended by an efficient set of 45° rotated features, which added additional domain-knowledge to the learning framework. Secondly, they derive a new post optimization procedure for a given boosted classifier that improves its performance significantly. Zhang, Ding and Liu (2015), also proposed an improved approach of PCA based on facial expression recognition algorithm using Fast Fourier Transform (FFT) during the preprocessing stage. They combined the amplitude spectrum of one image with phase spectrum of another image as a mixed image.

An important goal in image recognition is the ability to rate face recognition algorithms on the merit of efficiency and consistency in recognizing face images under variable environmental constraints. Until now, a face recognition algorithm's rate, runtime, sensitivity and descriptive statistics are the basic means of rating face recognition algorithms' performance. Delac and Grgic (2005) used some descriptive statistics to measure performance of face recognition algorithms. In their paper, they introduced measures of central tendencies, measures of dispersion, skewness and kurtosis of some template-based recognition algorithms and subsequently analysed the probability distribution of these algorithms. Beveridge *et al*., (2001) also investigated only Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in not as much detail using descriptive statistics.

This work focuses on statistical evaluation of the recognition performance of PCA/SVD and Whitened PCA/SVD under variable environmental constraints (variable facial expressions). This research explores and compares techniques for automatically recognizing facial actions in sequence of images or detecting an "unknown" human face in input imagery and recognizing the faces under various environmental constraints. This paper uses more intrinsic statistical methods (Multivariate methods) to assess the performance of face recognition algorithms under variable environmental constraints. The research methods, results, discussion and conclusions are presented in subsequent sections.

2. Methods

2.1 Data Acquisition

A real time face image database is created for the purpose of benchmarking the face recognition system. Two hundred and ninety four (42 individuals) labeled frontal facial images were randomly acquired from Cohn Kanade, Japanese Female Facial Expressions database (JAFFE) at labeled faces in the wild and some local Ghanaian students facial database. Of Two hundred and ninety four images, one hundred and eighty two facial images from 26 individuals were collected from the Cohn-Kanade AU-Coded Facial Expression Database along the seven universally accepted principal emotions (Neutral, Angry, Happy, Fear, Disgust, Sad, and Surprise). Subjects in the available portion of the database were 26 university students enrolled in introductory psychology classes. They ranged in age from 18 to 30 years. Forty two (6 individuals) images were also from the Local Ghanaian database. In the creation of the database, the observation room was equipped with a chair for the subject and one canon camera. Only image data from the frontal camera were captured. Subjects were instructed by an experimenter to perform a series of 7 facial displays that included single action units. Subject began and ended each display from a neutral face. Before performing each display, an experimenter described and modeled the desired display. Six of the displays were based on descriptions of prototypic basic emotions (happy, surprise, anger, fear, disgust, and sadness). Image sequences from neutral to target display were digitized into 256 by 256 or with 8-bit precision for grayscale values. Seventy frontal face images (10 individuals) were also collected from Japanese Female Facial Expressions database (JAFFE) along the principal emotional constraints. All three databases were combined in the study. This helped to evaluate the face recognition algorithms on large and different databases. The new created GFD accounted for the originality of the study database. The study database was divided into two subsets, training database and testing database. The training database comprised all 42 neutral poses and testing database comprised the remaining 210 expressions (Angry, Disgust, Fear, Happy, Sad and Surprise). Figure 1 shows a section of the study database.

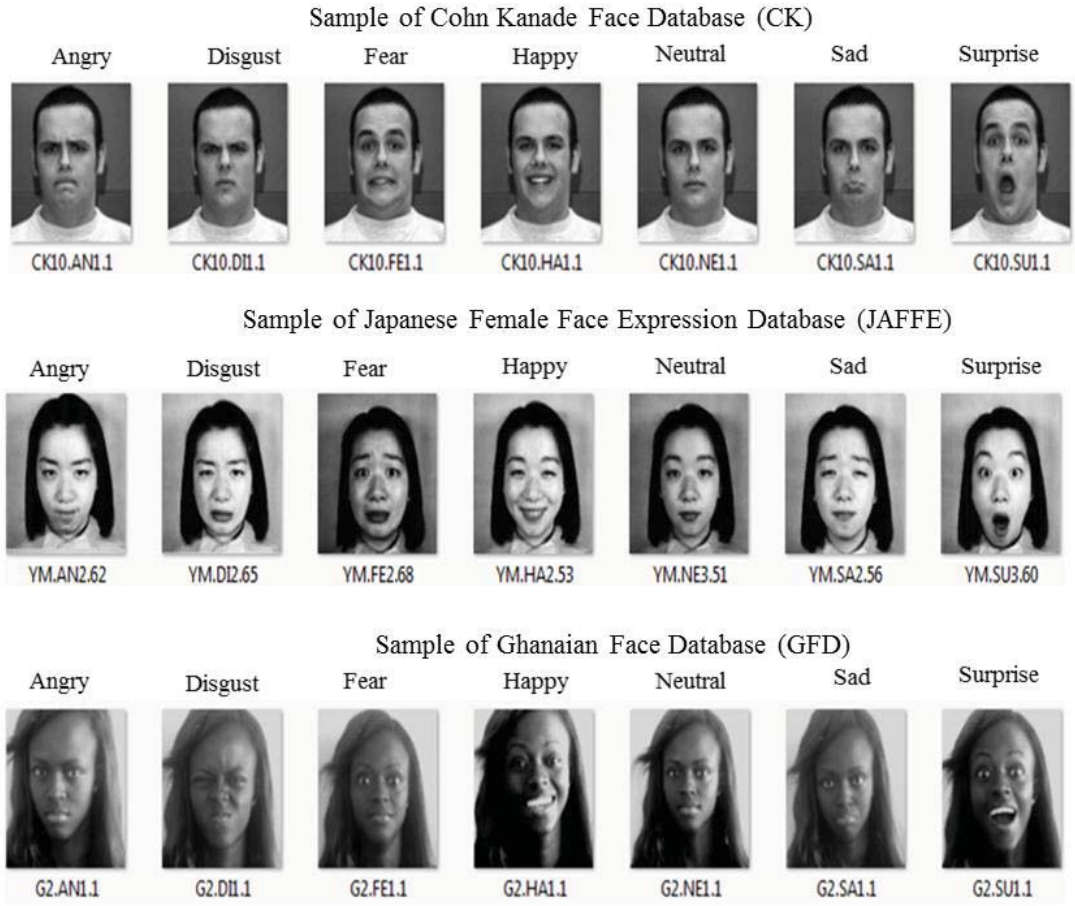


Figure 1. Sample of Research Database

2.2 Recognition Procedure

The study focused on running PCA/SVD and Whitened PCA/SVD recognition algorithms on a created face database. The research evaluated the recognition performance of the algorithms and subsequently compared their results on the created face database.

Face image data were passed to face recognition modules as input for the system. The face images passed were transformed into operationally compatible format (resizing images into uniform dimension). The data type of the image samples were also changed into double precision and passed for preprocessing. The entire recognition exercise comprises a preprocessing stage, feature extraction stage and recognition stage. The adopted preprocessing procedures are basically, mean centering and whitening. This is to help reduce the noise level and make the estimation process simpler and better conditioned.

The selected template based algorithms were used to train the created image database. In the extraction unit, unique face image features were extracted and stored for recognition. The obtained facial features were passed to the classifier unit for classification of a given face query with the knowledge created for the available database.

For the implementation of the facial recognition, a real time database was created. For the implementation of the proposed recognition design, the database samples were trained for the knowledge creation and classification. In the course of the training phase, when a new facial image was added to the system, the features were calculated according to a particular recognition algorithm's procedure and aligned for the dataset information. The test face weight and the known weight in the database are compared by finding the norm of the difference between the test and known weights. A maximum and minimum difference signifies poor and close match respectively. Figure 2 is a design of the entire face recognition process.

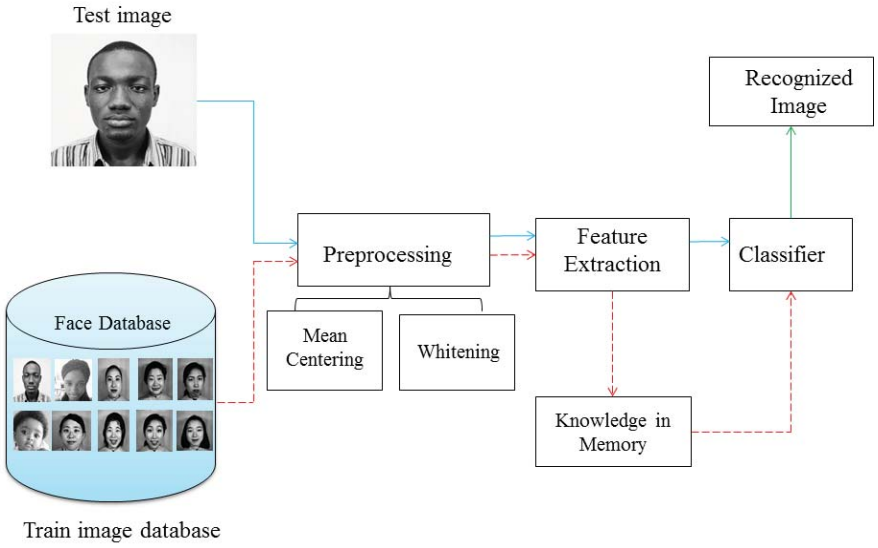


Figure 2. Research Design

2.3 Preprocessing of Frontal Face Image

Before applying any template-based algorithm on image data to be trained, it is useful to do some preprocessing. In this work, preprocessing is basically, Mean Centering and Whitening. This as indicated earlier on, is to help reduce the noise level and make the estimation process simpler and better conditioned.

As an illustration of preprocessing, Figure 3 shows six images selected from Japanese Female Face Expression database (JAFFE).

Figure 3. Six selected images from JAFFE

Define the image matrix, $\boldsymbol{M}_j$ as;

$$\boldsymbol{M}_j = (\boldsymbol{m_{jik}});\ i,k = 1,2,\ldots,p;\ j = 1,2,\ldots,n$$
$$= (\boldsymbol{m}_{ji1}, \boldsymbol{m}_{ji2},\ldots,\boldsymbol{m}_{jip}),$$

$$\boldsymbol{m_{jik}} = (m_{j1k}, m_{j2k}, \ldots, m_{jpk})^T, k = 1,2,\ldots,p$$

$$\boldsymbol{X}_j = (\boldsymbol{m}_{ji1}^{\boldsymbol{T}}, \boldsymbol{m}_{ji2}^{\boldsymbol{T}}, \ldots, \boldsymbol{m}_{jip}^{\boldsymbol{T}})^T \quad (1.0)$$

Where,

$p = the\ order\ of\ the\ image\ matrix.$

$n = the\ number\ of\ images\ to\ be\ trained.$

Now from equation (1.0), clearly, $\boldsymbol{X}_j$ is a column vector of dimension $N = p \times p$ and can be written as;

$$\boldsymbol{X}_j = \left(X_{ji}\right)_{N\times 1}, j = 1, 2, \ldots, n \tag{2.0}$$

where X_{ji} replaces the m_{jik} position wise.

The preprocessing steps are based on the sample $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ whose elements are the vectorised form of the individual images in the study.

2.3.1 Centering

This is a simple preprocessing step, executed by subtracting the mean,

$\bar{\boldsymbol{m}}_j = E(\boldsymbol{X}_j)$ of the data $(X_j, j = 1, 2, \ldots, n)$, from the data.

$$= \frac{1}{N}\sum_{i=1}^{N} X_{ji} \tag{3.0}$$

$$\bar{\boldsymbol{m}}_j = \frac{1}{N}\sum_{i=1}^{p}\sum_{k=1}^{p} \boldsymbol{m}_{jik,}\ (j = 1, 2, 3, \ldots, n)$$

where $N = (p \times p)$, $length\ (= rows\ of\ image \times columns\ of\ image)\ of\ the\ image\ data,\ X_j$.

Define $\bar{\boldsymbol{X}}_j$ as a constant column vector of order $(N = p \times p)$ with all elements same as $\bar{\boldsymbol{m}}_j$ $(j = 1, 2, \ldots, n)$.

The centered mean is denoted by, $\boldsymbol{W} = (w_1, w_2, w_3, \ldots, w_n)$; with

$$\boldsymbol{w_j} = \boldsymbol{X}_j - \bar{\boldsymbol{X}}_j,\ j = 1, 2, \ldots, n \tag{4.0}$$

Applying equation (4.0) to the images in Figure 2, the generated mean centered images are shown in Figure 4.

Figure 4. Six mean centered images from JAFFE

2.3.2 Whitening

Whitening is a preprocessing technique that removes the noise factors in the observed image data, $\boldsymbol{X}$ so as to obtain a new image, $\vec{\boldsymbol{X}}$ with uncorrelated components but equal unit variance. This is to say, the covariance of $\vec{\boldsymbol{X}}$ is the identity matrix, $\boldsymbol{I}$. A simple way to whiten images is to find the eigenvectors and eigenvalues of the observed images through eigenvalue decomposition (for symmetric image matrix) or singular value decomposition (for asymmetric image matrix) of the covariance matrix. Suppose the covariance matrix, $\boldsymbol{C}$ is given by;

$$\boldsymbol{C} = \frac{1}{n}\boldsymbol{W}^T\boldsymbol{W} \tag{5.0}$$

Define matrix $\boldsymbol{U} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots \boldsymbol{u}_n)$ where $\boldsymbol{u_j}, \boldsymbol{j} = 1, 2, \ldots, n$ is the jth eigenvector of the covariance matrix $\boldsymbol{C}$. Let $\boldsymbol{D}$ $(dimension\ n \times n)$ be the diagonal matrix whose entries $(\lambda_{jj}, \boldsymbol{j} = 1, 2, \ldots, n)$ are the eigenvalues corresponding to the eigenvectors $\boldsymbol{u}_j, j = 1,2, \ldots, n$. The whitened images $\vec{\boldsymbol{X}}$ are given by;

$$\vec{X} = UD^{-\frac{1}{2}}U^T X^T \tag{6.0}$$

The covariance matrix $\vec{C}$ of $\vec{X}$ is given by;

$$\vec{C} = \frac{1}{n}\vec{X}^T\vec{X} = I \tag{7.0}$$

Figure 5 shows the whitened outcome of the six images shown in Figure 3.

Figure 5. Whitened images

The whitened matrix, $\vec{X}$, built from the eigenvalue decomposition of the covariance matrix C of the zero-mean observation, w, creates a set of uncorrelated unit image variables.

2.4 Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) transforms correlated variables into a set of uncorrelated ones that better expose the various relationships among the original data items, while at the same time, identifying and ordering the dimensions along which data points exhibit the most variation. Once SVD has identified the most variation, it is possible to find the best approximation of the original data points using fewer dimensions (Baker, 2005). Hence, SVD can be seen as a method for data reduction or dimensionality reduction. Consider an arbitrary real $m \times n$ matrix A, then there are orthogonal matrices U and V and a diagonal matrix Σ, such that, $A = U\Sigma V^T$, where U is an $m \times n$ matrix, V is an $n \times n$ matrix and Σ is an $n \times n$ diagonal matrix with diagonal entries $\sigma_{ii} \geq 0$, $\forall i = 1, 2, \ldots, n$ and $\sigma_{11} \geq \sigma_{22} \geq \cdots \geq \sigma_{nn}$. In practice, the components of Σ are unknown and are to be estimated. The columns of U and V are called the singular vectors corresponding to the positive values (singular values) in the diagonal matrix Σ. When these are used to represent vectors in the domain and range of transformation, the transformation simply dilates and contracts some components according to the magnitude of the singular values and possibly discards values and appends zeros as needed to account for a change in dimension. It is therefore clear that SVD tells how to choose orthonormal bases so that the transformation is represented by a matrix with the simplest possible form.

2.5 Principal Component Analysis (PCA)

PCA is concerned with elucidating the covariance structure of a set of variables. It seeks to find a set of basis images which are uncorrelated, that is, they cannot be linearly predicted from each other and also yield projection directions that maximize the total scatter across all classes or across all face images. According to Barlett *et al.,* (2002), PCA can thus be seen as partially implementing Barlow's ideas: Dependencies that show up in the joint distribution of pixels are separated out into marginal distribution of PCA coefficients. Most of the successful representations for face recognition, such as eigenface and local feature analysis are based on PCA.

2.6 Feature Extraction

Having these algorithms in mind, it is now time to seek a set of n orthonormal vectors, e_j, which best describes the distribution of the data. The t^{th} vector e_t is chosen such that

$$\lambda_t = \frac{1}{n}\sum_{j=1}^{n}\left(e_j^T w_j\right)^2$$

is maximum subject to the orthonormality constraints.

$$e_j^T e_t = \delta_{lt} = \begin{cases} 1 & if, l = t \\ 0, & elsewhere \end{cases}$$

The vectors $\boldsymbol{e_t}$ and scalars λ_t are the eigenvectors and eigenvalues respectively of the covariance matrix $\boldsymbol{C}$.

The size of $\boldsymbol{C}$ $(n \times n)$ could be enormous and determining the eigenvectors and eigenvalues is an intractable task for typical image sizes. A known theorem in linear algebra states that the vectors $\boldsymbol{e_j}$ and the scalars $\boldsymbol{\lambda_j}$ can be obtained by solving for the eigenvalues of $\boldsymbol{W^T W}$, respectively.

$$\boldsymbol{WW^T d_j} = \boldsymbol{u_j}\,(\boldsymbol{W d_j}) \tag{8.0}$$

This means that the first $n-1$ eigenvectors, $\boldsymbol{e_j}$, and eigenvalues, λ_j, of $\boldsymbol{WW^T}$ are given by $\boldsymbol{W d_j}$ and $\boldsymbol{u_j}$ respectively. $\boldsymbol{W d_j}$ needs to be normalized in order to be equal to $\boldsymbol{e_j}$.

Hence, $\boldsymbol{e_j} = \sum_{j=1}^{n} \boldsymbol{w_j d_j}$ where $\boldsymbol{d_j}$ and $\boldsymbol{w_j}$ are the columns from $\boldsymbol{U}$ and $\boldsymbol{W}$ respectively.The principal components of the trained image set are determined by computing;

$$\boldsymbol{v_j} = \boldsymbol{e_j^T}\left(\boldsymbol{X_j} - \overline{\boldsymbol{m}}\right), \tag{9.0}$$

where $\boldsymbol{\Omega} = [\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}]$

The large correlated image dimensions are finally reduced to uncorrelated smaller intrinsic dimensions which display important characteristics of the image set. An unknown input face is passed through the steps below before identification.

Following the steps in the feature extraction stage, a new face from the test image database is transformed into its eigenface components. First the input image is compared with the mean image (trained images mean) in memory and their difference is multiplied with each eigenvector from $\boldsymbol{e_j}$. Each value represents a weight and is saved on a vector $\boldsymbol{\Omega}$. This is done by looking for the face class that minimizes the Euclidean; $\epsilon_r = |\boldsymbol{\Omega} - \boldsymbol{\Omega_r}|$. Figure 6 is a flow diagram of the study algorithms.

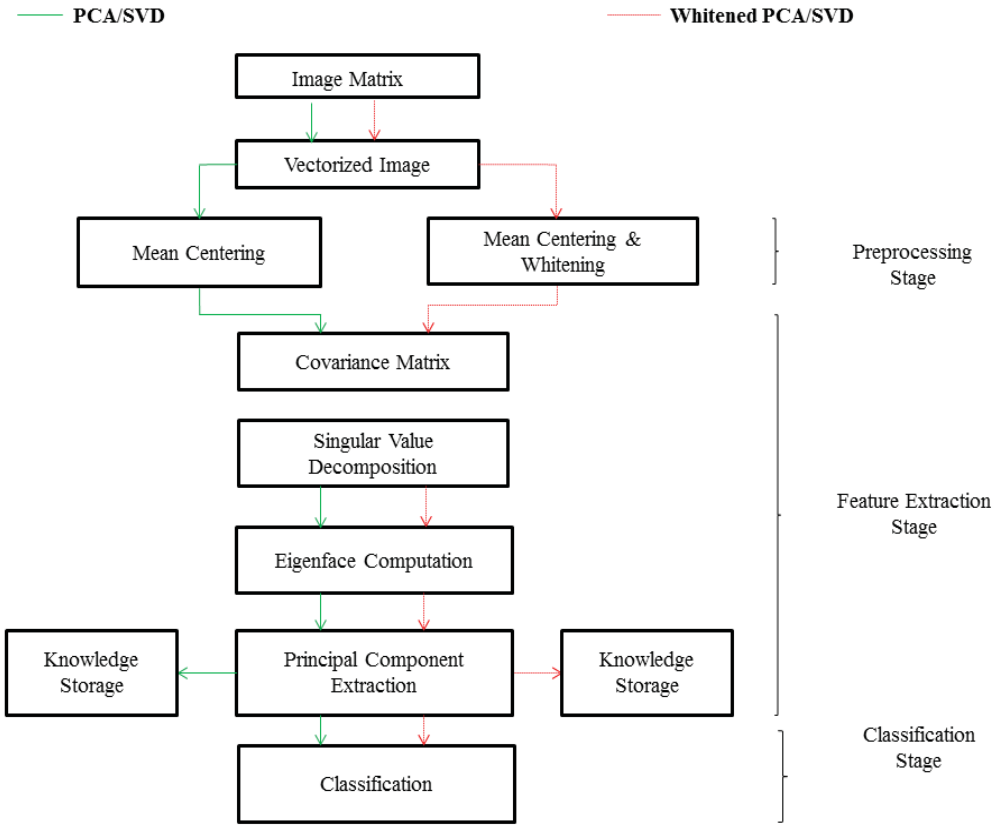


Figure 6. Flow diagram of study algorithms.

2.7 Results and Discussion

This section presents the statistical procedures used to evaluate the fore-mentioned recognition algorithms. The results of running these statistical tests on the study dataset are also presented and discussed.

2.7.1 Statistical Evaluation of the Face Recognition Algorithms

The recognition algorithms under study are PCA and SVD with Mean Centering as the preprocessing step (Algorithm 1) and PCA and SVD with Mean Centering and Whitening as the preprocessing step (Algorithm 2).

From the study database, 6-variates are collected per each algorithm from the Euclidean distance between the universally accepted principal emotions (Angry, Disgust, Fear, Happy, Sad and Surprise) and their neutral pose. (see *appendix 2.0* for data).

In assessing multivariate normality, a chi-square plot of the datasets (Algorithm specific) is done by plotting the generalized squared distances of the datasets against the chi-square quantiles.

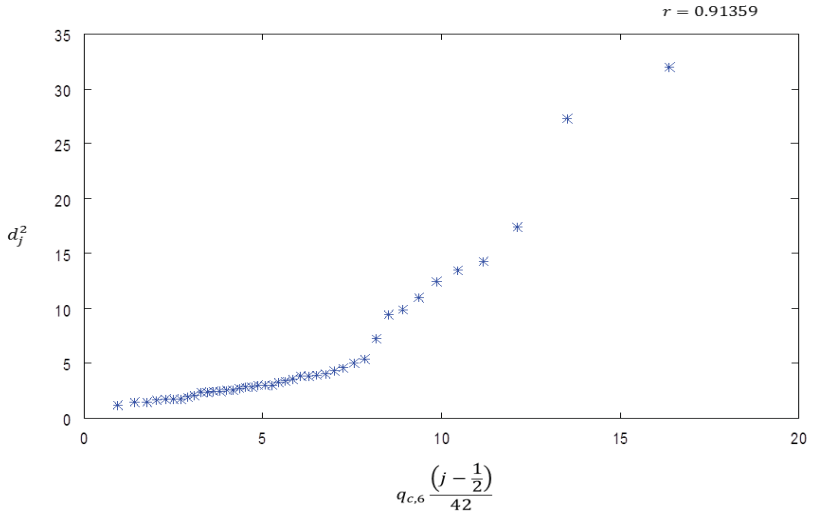


Figure 7. Chi-square plot of Algorithm 1

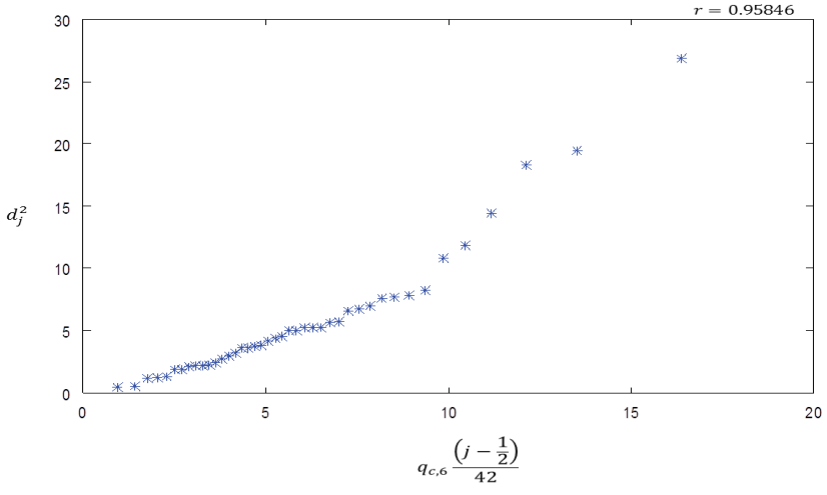


Figure 8. Chi-square plot of Algorithm 2

Figure 7 and Figure 8 show the chi-square plots of the datasets from the study algorithm 1 and Algorithm 2 respectively. The correlation, r, values are 0.91359 and 0.95846 for algorithm 1 and algorithm 2 respectively are close to 1. These satisfy the assumption of a unit slope of the chi-square plot. Multivariate normality exists and hence can be assumed in subsequent statistical test that will be performed on the datasets.

2.7.2 Repeated Measures Design

The purpose of the test is to determine whether for each of the recognition algorithms under study, there exist

significant differences between the average distances of the various poses from their neutral pose.

Using the 6-variate dataset from Algorithm 1, we have;

$$n = 42, C\overline{X} = \begin{bmatrix} 1003.917 \\ -98.952 \\ 278.283 \\ -96.397 \\ -1167.166 \end{bmatrix}$$

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

$$(C\Sigma C^T)^{-1} = \begin{bmatrix} 1.28e^{-007} & 1.12e^{-007} & 6.21e^{-008} & -3.69e^{-009} & 6.48e^{-009} \\ 1.12e^{-007} & 4.99e^{-007} & 4.15e^{-007} & 1.33e^{-007} & 1.40e^{-008} \\ 6.21e^{-008} & 4.15e^{-007} & 7.85e^{-007} & 5.68e^{-007} & 2.40e^{-007} \\ -3.69e^{-009} & 1.33e^{-007} & 5.68e^{-007} & 7.40e^{-007} & 4.36e^{-007} \\ 6.48e^{-009} & 1.40e^{-008} & 2.40e^{-007} & 4.36e^{-007} & 5.44e^{-007} \end{bmatrix}$$

Hence, the $T^2 -$ statistic;

$$T^2 = \boldsymbol{n}\big(\boldsymbol{C\overline{X}}\big)(\boldsymbol{C\Sigma C^T})^{-1}\boldsymbol{C\overline{X}}$$

$$= 35.095$$

Now,

$$\frac{(n-1)(p-1)}{(n-p+1)} = 5.540541, \quad F_{p-1,n-p+1}(0.05) = 2.4697$$

$$\frac{(n-1)(p-1)}{(n-p+1)} F_{p-1,n-p+1}(0.05) = 13.6832$$

Reject $H_0 : \boldsymbol{C\mu} = \boldsymbol{0}$ if;

$$T^2 = \boldsymbol{n}\big(\boldsymbol{C\overline{X}}\big)(\boldsymbol{C\Sigma C^T})^{-1}\boldsymbol{C\overline{X}} > \frac{(n-1)(p-1)}{(n-p+1)} F_{p-1,n-p+1}(\alpha)$$

$35.095 > 13.6832$. There is therefore enough evidence at 5% level of significance to reject H_0 and conclude on $H_1 : \boldsymbol{C\mu} \neq \boldsymbol{0}$. This means there exist significant difference in the average distances of the various constraints from their neutral pose when Algorithm 1 is used for recognition.

Using the 6-variate dataset from Algorithm 2, we have;

$$n = 42, C\overline{X} = \begin{bmatrix} 440.23 \\ -271.21 \\ 865.49 \\ -157.19 \\ -2273.87 \end{bmatrix}$$

$$(C\Sigma C^T)^{-1} = \begin{bmatrix} 2.51e^{-007} & 8.90e^{-008} & 2.73e^{-008} & 7.67e^{-008} & 1.05e^{-007} \\ 8.90e^{-008} & 2.04e^{-007} & 1.47e^{-007} & 1.35e^{-007} & 4.95e^{-008} \\ 2.73e^{-008} & 1.47e^{-007} & 2.79e^{-007} & 1.53e^{-007} & -7.36e^{-009} \\ 7.67e^{-008} & 1.36e^{-007} & 1.53e^{-007} & 2.16e^{-007} & 7.32e^{-007} \\ 1.05e^{-007} & 4.95e^{-008} & -7.36e^{-009} & 7.32e^{-007} & 2.19e^{-007} \end{bmatrix}$$

Hence, the $T^2 -$ statistic;

$$T^2 = \boldsymbol{n}\left(\boldsymbol{C}\overline{\boldsymbol{X}}\right)(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^{\boldsymbol{T}})^{-1}\boldsymbol{C}\overline{\boldsymbol{X}}$$

$$= 51.645$$

Now,

$$\frac{(n-1)(p-1)}{(n-p+1)} = 5.540541, \quad F_{p-1,n-p+1}(0.05) = 2.4697$$

$$\frac{(n-1)(p-1)}{(n-p+1)} F_{p-1,n-p+1}(0.05) = 13.6832$$

$51.645 > 13.6832$. There is therefore enough evidence at 5% level of significance to reject H_0 and conclude on $H_1: \boldsymbol{C\mu} \neq \boldsymbol{0}$. This means there exist significant difference in the average distances of the various constraints from their neutral pose when Algorithm 2 is used for recognition. The 95% simultaneous confidence intervals for the estimates of the mean differences are shown in Table 1.

Table 1. Simultaneous Confidence Intervals.

	Algorithm 1		Algorithm 2	
Constraints	Lower	Upper	Lower	Upper
Angry vs Disgust	−810.2300	2818.0636	−922.7925	1803.2427
Angry vs Fear	−949.4138	2759.3435	−1805.9279	2143.9577
Angry vs Happy	−537.6334	2904.1299	−1114.9863	3184.0031
Angry vs Sad	−879.9624	3053.6657	−828.3633	2582.9919
Angry vs Surprise	−1874.7483	1714.1202	−3145.0346	351.933
Disgust vs Fear	−1489.3901	1291.4861	−2150.1650	1607.7446
Disgust vs Happy	−925.4309	1284.0937	−1321.6768	2510.2434
Disgust vs Sad	−1201.0698	1366.9394	−1083.5872	1957.7656
Disgust vs Surprise	−2317.2119	148.7501	−3780.7245	106.9097
Fear vs Happy	−1290.6203	1847.1871	−739.2223	2470.2092
Fear vs Sad	−855.2611	1219.0347	−1109.0879	2525.6867
Fear vs Surprise	−2024.8504	54.2926	−3808.7245	677.593
Happy vs Sad	−1633.8636	1441.0704	−2038.4331	1724.045
Happy vs Surprise	−2469.2859	−57.8386	−4401.5625	−460.5558
Sad vs Surprise	−2298.1676	−36.1638	−3745.1675	−802.5626

2.6.2 Paired Comparisons

Measurements are often recorded under different sets of experimental conditions to see whether the responses differ significantly over these sets. In the case of this study, the Euclidean norms of various poses (Angry, Disgust, Fear, Happy, Sad and Surprise) along with their neutral pose are recorded by using two different recognition algorithms. Specifically for this study, 42 individuals were tested on the different recognition algorithms. The paired responses are analyzed by computing their differences, thereby eliminating much of the influence of extraneous unit to unit variation.

The multivariate case is motivated for 6 constraints, 2 algorithms and 42 experimental units. The paired differences random variables are;

$$\boldsymbol{D}_j^T = [D_{j1}, D_{j2}, \dots, D_{j6}] \text{ for } j = 1,2, \dots ,42$$

$$\overline{D} = \frac{1}{42}\sum_{j=1}^{42} D_j = \begin{bmatrix} -896.43 \\ -2345.95 \\ -2369.30 \\ -2092.28 \\ -3060.31 \\ -2034.09 \end{bmatrix}$$

$$\text{and} \quad \boldsymbol{\Sigma}_d = \frac{1}{42-1}\sum_{j=1}^{42}\left(\boldsymbol{D}_j - \bar{\boldsymbol{D}}\right)\left(\boldsymbol{D}_j - \bar{\boldsymbol{D}}\right)^T$$

$$= \begin{bmatrix} 1.91e^{+007} & 9.05e^{+006} & 3.26e^{+006} & 6.54e^{+006} & 7.94e^{+006} & 7.61e^{+006} \\ 9.05e^{+006} & 1.66e^{+007} & 5.38e^{+006} & 1.26e^{+007} & 9.51e^{+006} & 7.99e^{+006} \\ 3.26e^{+006} & 5.38e^{+006} & 1.04e^{+007} & 7.33e^{+006} & 7.18e^{+006} & 3.23e^{+006} \\ 6.54e^{+006} & 1.26e^{+007} & 7.33e^{+006} & 2.04e^{+007} & 8.10e^{+006} & 9.54e^{+006} \\ 7.94e^{+006} & 9.51e^{+006} & 7.18e^{+006} & 8.10e^{+006} & 1.34e^{+007} & 9.35e^{+006} \\ 7.61e^{+006} & 7.99e^{+006} & 3.23e^{+006} & 9.54e^{+006} & 9.35e^{+006} & 1.47e^{+007} \end{bmatrix}$$

$$\boldsymbol{\Sigma}_d = \begin{bmatrix} 7.89e^{-008} & -3.10e^{-008} & 4.54e^{-009} & 8.32e^{-009} & -1.97e^{-008} & -1.79e^{-008} \\ -3.10e^{-008} & 1.65e^{-007} & 3.23e^{-008} & -8.17e^{-008} & -8.43e^{-008} & 2.62e^{-008} \\ 4.54e^{-009} & 3.23e^{-008} & 2.03e^{-007} & -7.19e^{-008} & -1.40e^{-007} & 7.15e^{-008} \\ 8.32e^{-009} & -8.17e^{-008} & -7.19e^{-008} & 1.30e^{-007} & 5.89e^{-008} & -6.59e^{-008} \\ -1.97e^{-008} & -8.43e^{-008} & -1.40e^{-007} & 5.89e^{-008} & 2.72e^{-007} & -1.24e^{-007} \\ -1.79e^{-008} & 2.62e^{-008} & 7.15e^{-008} & -6.59e^{-008} & -1.24e^{-007} & 1.69e^{-007} \end{bmatrix}$$

In addition, $\boldsymbol{D}_j \ , j = 1, 2, \ldots, 42$ are independent individuals, $N_6\,(\mu, \boldsymbol{\Sigma}_d)$. Given the observed difference, $\boldsymbol{d}_j^T = [d_{j1}, d_{j2}, \ldots, d_{j6}], j = 1, 2, \ldots, 42$

$$T^2 = n\,\bar{\boldsymbol{d}}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\mathrm{d}}^{-\mathbf{1}}\bar{\boldsymbol{d}} = 35.742$$

$$\left[\frac{(n-1)p}{(n-p)}\right]F_{p,n-p}(\alpha)$$

$$= \left[\frac{(42-1)6}{(42-6)}\right]F_{6,42-6}(0.05) = 16.152$$

$35.742 > 16.152$. This assertion of equal mean difference between the algorithms is not tenable at 5% level of significance. It can be concluded that, there exist significant difference in the average distances of both algorithms with respect to the study constraints (pose-wise).

The Bonferroni 95% simultaneous confidence intervals for the individual mean difference is given by;

$\mu_i: \bar{d}_i \pm t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{S_{d_i}^2}{n}}$, where $\bar{d}_i$ is the ith element of $\bar{\boldsymbol{D}}$, $S_{d_i}^2$ is the ith diagonal of $\boldsymbol{\Sigma}_d$ and

$t_{n-1}\left(\frac{\alpha}{2p}\right)$ is the upper $100\left(\frac{\alpha}{2p}\right)th$ percentile of the t-distribution.

These confidence intervals will reveal specifically which constraints have significant differences in Euclidean distances when the different face recognition algorithms are used. Table 2 below shows the confidence intervals of estimates for the average of the difference in distances.

Table 2. Bonferroni Simultaneous Confidence Intervals

Constraints	Average differences	Lower	Upper
Angry poses	$\bar{d}_1 = -896.43$	-2764.2480	971.3904
Disgust poses	$\bar{d}_2 = -2345.95$	-4086.8883	-605.0192
Fear poses	$\bar{d}_3 = -2369.3$	-3750.5892	-988.0069
Happy poses	$\bar{d}_4 = -2092.28$	-4025.6741	-158.8898
Sad poses	$\bar{d}_5 = -3060.31$	-4628.7849	-1491.8397
Surprise poses	$\bar{d}_6 = -2034.09$	-3673.1152	-395.0652

This means for the two algorithms (Algorithm 1 and Algorithm 2) under study, there exist significant difference in their poses (Disgust, Fear, Happy, Sad and Surprise) recognition except their recognition of the angry pose. $d_1 = [-2764.2480, 971.3904]$ means, there is no significant difference in the average recognition distance on Angry pose between Algorithm 1 and Algorithm 2. It can therefore be inferred that, at 5% level of

significance, both algorithms have significantly different average recognition distances for all poses except angry pose.

2.7.3 Test of Equality of Covariance Matrices (Box's M-Test)

This test will be used as a measure of consistency between the recognition algorithms. The test will reveal whether the variations in distances across algorithms in recognizing face images in the study database are equal or significantly different. The most consistent algorithm should have lower variation in recognition distances. The Box's test is based on the χ^2 approximation to the sampling distribution of M.

$$M = \sum_{l=1}^{g}(n_l - 1)\ln|\boldsymbol{S}_{pooled}| - \sum_{l=1}^{g}(n_l - 1)\ln|\boldsymbol{S}_l|, \quad g = 2\ (number\ of\ algorithms)$$

From the data, $|\boldsymbol{S}_1| = 5.6378e^{+038}, |\boldsymbol{S}_2| = 3.8460e^{+040}, |\boldsymbol{S}_{pooled}| = 4.9885e^{+051}$

$$M = 2271.4$$

$$U = \left[\sum_{l=1}^{g}\frac{1}{(n_l - 1)} - \frac{1}{\sum_{l=1}^{g}(n_l - 1)}\right]\left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)}\right]$$

$$= 0.077526$$

where $p = 6$ is the number of constraints.

Now,

$$K = (1 - U)M$$

$$= (1 - 0.077526)2271.4$$

$$= 2095.3$$

$$\chi^2_{\frac{1}{2}p(p+1)(g-1)}(0.05) = 32.671$$

$2095.3 > 32.671$, hence the assertion of equality of covariance is not tenable at 5% level of significance. We can therefore conclude that, the covariance of Algorithm 1 and Algorithm 2 are not equal. This means, the variations in the Algorithm 1 and Algorithm 2 recognition distances are significantly different.

2.7.4 Profile Analysis

For small sample size, profile analysis depends on the normality assumption (Johnson, & Wichern, 2007). The datasets under study are multivariate normal; hence this assumption of normality is satisfied. Profile analysis also works on the premise of equality of covariance matrices. Here, the pooled covariance is then used as the common covariance for the populations under study. The Box's M test revealed that, the covariance matrices of the algorithms under study are unequal. According Mettle, Yeboah and Asiedu (2014), the profile analysis is still feasible when the assertion of equality of covariance matrix is not tenable. That is, profile analysis can continue when unequal covariance exist. In this case the separate covariance matrices are used in the computation.

Now from the study datasets,

$$(\bar{\boldsymbol{X}}_2 - \bar{\boldsymbol{X}}_1)^T = [896.43\ \ 2345.95\ \ 2092.28\ \ 3060.31\ \ 2034.09]$$

$$\boldsymbol{S}_1 = \begin{bmatrix} 1.10e^{+007} & 2.67e^{+006} & 8.71e^{+005} & 3.87e^{+006} & 7.66e^{+005} & 1.71e^{+006} \\ 2.67e^{+006} & 4.47e^{+006} & -6.72e^{+004} & 3.29e^{+006} & 9.24e^{+005} & 1.07e^{+006} \\ 8.71e^{+005} & -6.72e^{+004} & 1.33e^{+006} & -1.82e^{+005} & 2.34e^{+005} & 1.70e^{+005} \\ 3.87e^{+006} & 3.29e^{+006} & -1.82e^{+005} & 5.86e^{+006} & 5.23e^{+005} & 1.86e^{+006} \\ 7.66e^{+005} & 9.24e^{+005} & 2.34e^{+005} & 5.23e^{+005} & 2.44e^{+006} & 4.20e^{+005} \\ 1.71e^{+006} & 1.07e^{+006} & 1.70e^{+005} & 1.86e^{+006} & 4.20e^{+005} & 2.33e^{+006} \end{bmatrix}$$

$$\boldsymbol{S}_2 = \begin{bmatrix} 6.79e^{+006} & 5.94e^{+006} & 1.91e^{+006} & 2.55e^{+006} & 4.27e^{+006} & 3.82e^{+006} \\ 5.94e^{+006} & 1.08e^{+007} & 4.48e^{+006} & 6.00e^{+006} & 7.19e^{+006} & 4.72e^{+006} \\ 1.91e^{+006} & 4.48e^{+006} & 8.99e^{+006} & 6.78e^{+006} & 4.77e^{+006} & 1.89e^{+005} \\ 2.55e^{+006} & 6.00e^{+006} & 6.78e^{+006} & 1.25e^{+007} & 6.15e^{+006} & 5.40e^{+006} \\ 4.27e^{+006} & 7.19e^{+006} & 4.77e^{+006} & 6.15e^{+006} & 1.07e^{+007} & 7.14e^{+006} \\ 3.82e^{+006} & 4.72e^{+006} & 1.89e^{+005} & 5.40e^{+006} & 7.14e^{+006} & 1.02e^{+007} \end{bmatrix}$$

The sample sizes, $n_1 = n_2 = 42$

$$T^2 = (\bar{X}_2 - \bar{X}_1)^T C^T \left[C \left(\frac{S_1}{n_1} + \frac{S_2}{n_2} \right) C^T \right] C(\bar{X}_2 - \bar{X}_1)$$

$$= 10.459$$

$$c^2 = \left[\frac{(n_1 + n_2 - 2)(p-1)}{n_1 + n_2 - p} \right] F_{p-1, n_1+n_2-p}(0.05)$$

$$\left[\frac{(42 + 42 - 2)(6-1)}{42 + 42 - 6} \right] F_{5,78}(0.05) = 2.451$$

10.459 > 2.451 and the assertion of parallel profiles is not tenable at 5% significance level. It can therefore be concluded that, the profiles of Algorithm 1 and Algorithm 2 are not parallel. This also implies that, the profiles are not coincident and subsequently not level. Figure 9 shows a mean plot of the recognition algorithms.

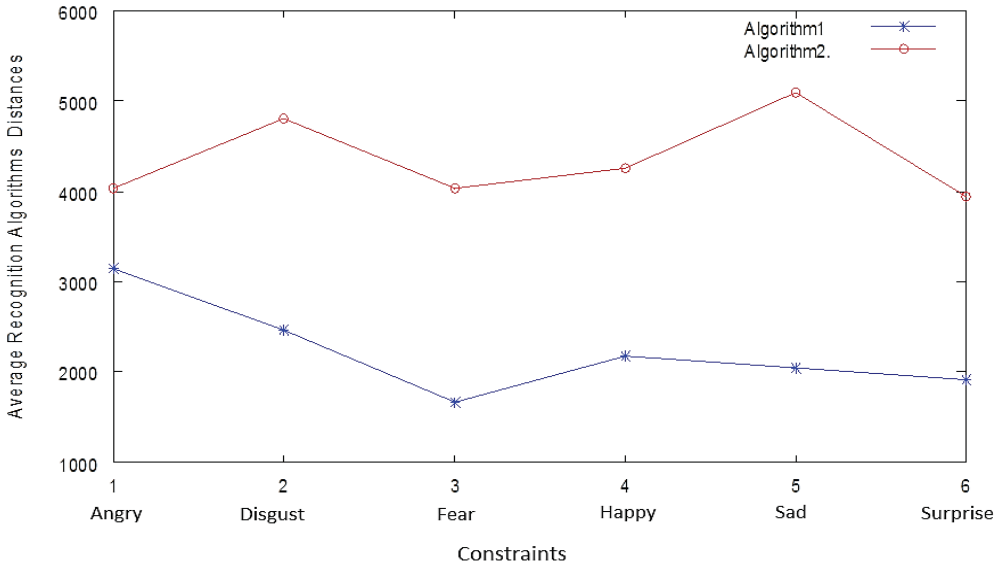


Figure 9. Mean plot of recognition distances.

2.7.5 Levene's Test of Equality of Variance

The goal of this test is to determine whether the Algorithms under study have equal variance in their recognition of the study constraints. The test is quite sensitive to the underlying assumption that the, samples been tested should come from a normal population.

In this study, two independent normal populations each from the different study algorithms are collected. For example, angry pose data from algorithm 1 tested against angry pose data from algorithm 2.

Let X_{jk1}, $j = 1, 2, \ldots, 42$ (individuals) and $k = 1, 2, \ldots, 6$ (poses) be the datasets from algorithm 1 and X_{jk2}, $j = 1, 2, \ldots, 42$ (individuals) and $k = 1, 2, \ldots, 6$ (poses) be the datasets from algorithm 2. Now consider two independent normal populations X_{jk1} and X_{jk2}, with unknown variance.

With samples of size $n_{k1} = 42$, from Algorithm 1, $n_{k2} = 42$, from Algorithm 2 and their respective sample variance S_{k1}^2 and S_{k2}^2.

A 95% confidence interval is given by;

$$\frac{S_{k1}^2}{S_{k2}^2} F_{1-\frac{\alpha}{2}, n_{2k-1}, n_{1k-1}} < \frac{\sigma_{k1}^2}{\sigma_{k2}^2} < \frac{S_{k1}^2}{S_{k2}^2} F_{\frac{\alpha}{2}, n_{2k-1}, n_{1k-1}}$$

$$F_{0.975,41,41} = 0.5375,\ F_{0.025,41,41} = 1.8604$$

The estimates of the ratio of variance are given by;

$$\left[\frac{\hat{S}_{11}^2}{\hat{S}_{12}^2} \quad \frac{\hat{S}_{21}^2}{\hat{S}_{22}^2} \quad \frac{\hat{S}_{31}^2}{\hat{S}_{32}^2} \quad \frac{\hat{S}_{41}^2}{\hat{S}_{42}^2} \quad \frac{\hat{S}_{51}^2}{\hat{S}_{52}^2} \quad \frac{\hat{S}_{61}^2}{\hat{S}_{62}^2}\right] = [1.6142 \quad 0.4141 \quad 0.1479 \quad 0.4697 \quad 0.2283 \quad 0.2274]$$

The 95% confidence intervals for the estimates of the ratio of variances are shown in Table 3 below.

Table 3. Confidence interval for the ration of variance

Constraints	Ratio of Variances	Lower	Upper
Angry poses	1.6142	0.8676	3.0029
Disgust poses	0.4141	0.2226	0.7704
Fear poses	0.1479	0.0795	0.2752
Happy poses	0.4697	0.2525	0.8738
Sad poses	0.2283	0.1227	0.4247
Surprise poses	0.2274	0.1222	0.4230

Clearly from Table 3, the confidence interval for angry poses $[0.8676, 3.0029]$ contains 1 and hence the assertion of equality of variance of the two algorithms is tenable at 5% significance level. The remaining constraints (Disgust, Fear, Happy, Sad and Surprise) have confidence intervals that do not contain 1. Here, assertion of equality of variance is not tenable. This means the variances of the recognition distances for these poses are not equal. Now considering the constraint for which equality of variance is not tenable (Disgust, Fear, Happy, Surprise and Sad), estimate of the ratio of variance are given as 0.42141, 0.1479, 0.4697, 0.2283 and 0.2274 respectively. All these ratios are less than 1 and hence we can reach the conclusion that, the variations in Algorithm 2 are greater than that of Algorithm 1 in the recognition of these constraints. Subsequently, Algorithm 1 is considered as comparatively consistent in the recognition of Disgust, Fear, Happy, Sad and Surprise poses.

3. Conclusion

The runtime of Algorithm 1 and Algorithm 2 in the recognition of the 252 images is 70.470 seconds and 191.79 seconds respectively. The time used by algorithm 2 in the whitening process accounts for the differences in the algorithms' runtime (speed). The recognition rates of Algorithm 1 and Algorithm 2 are 92.86% and 88.10% respectively. It is evident from the above statistical methods that, the algorithms considered are significantly different in recognizing all poses except the angry pose. Although both algorithms are equally consistent in recognizing angry pose, Algorithm 1 (PCA with SVD and mean centering as preprocessing step) is comparatively efficient (from recognition rate) and consistent (from variation) in recognizing all other constraints under study. Algorithm 1 is therefore adjudged as comparatively better in recognizing face images under the variable environmental constraints.

References

Agrawal, S., Khatri, P., & Gupta, S., (2014). Facial Expression Recognition Techniques: A survey. *ITM University. Gwalior, India.*

Baker, K. (2005). Singular value decomposition tutorial, Ohio.

Barlett, M. S., Movellan, J., & Sejnowski, T. J. (2002). Face recognition by independent component analysis: *IEEE Trans. on Neural networks*, *13*(6), 1450-1464. http://dx.doi.org/10.1109/TNN.2002.804287

Beveridge, J.R., She, K., Draper, B., & Givens, G. (2001). Parametric and Non-parametric Methods for the Statistical Evaluation of Human ID Algorithms. *IEEE Third Workshop on Empirical Evaluation Methods in Computer Vision*, Kauai, HI, USA,

Chiara, T., Viola Macchi Cassia, F. S., & Leo, I. (2006). Newborns face recognition: Role of inner and outer facial features. *Child Development, 77* (2), 297-311.

Delac, K., Grgic, M., & Grgic, S., (2005). Statistics in face Recognition: Analyzing probability distributions of PCA, ICA and LDA performance. *Proceedings 4th international symposium on image and signal processing and Analysis. ISPA.* pp. 289-294. http://dx.doi.org/10.1109/ISPA.2005.195425

Johnson, A. R., & Wichern, W. D., (2007). *Applied Multivariate Statistical Analysis. 6th Ed*. New Jersey: Pearson Prentice Hall

Lucey, P.,Cohn, J.F., Kanade,T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. *Proceedings of the 3rd International Workshop on CVP for Human Communicative Behavior Analysis (CVPR4HB) San Fran- cisco,* USA, 94-10. http://dx.doi.org/10.1109/cvprw.2010.5543262

Lienhart, R., & Maydt, J. (2002). An extended set of haar-like features for rapid object detection. *IEEE ICIP 2002, 1,* pp. 900-903. http://dx.doi.org/10.1109/icip.2002.1038171

Zhang, D., Ding, D., Li, J., & Liu, Q. (2015). PCA Based Extracting Feature Using Fast Fourier Transform for Facial Expression Regression. *Transaction on Engineering Technologies,* pp. 413 – 424. Springer Netherlands.

Marqués, I., (2010). Face Recognition Algorithms; Proyecto Fin de 1-5.

Mettle, F. O., Yeboah, E., & Asiedu, L. (2014). Profile Analysis of Spatial Differential of Inflation in Ghana. *International Journal of Statistics and Analysis*, *4*(2), 245-259.

Rahman, M. U., (2013). A comparative study on face recognition techniques and neural network, vol abs/1210-1916.

Voila, P., & Jones, M.J., (2001). Rapid Object Detection using a Boosted Cascade of Simple Feature. *IEEE CVPR.* http://dx.doi.org/10.1109/cvpr.2001.990517

Wagner, P. (2012). Face recognition with python. Retrieved from http://www.bytefish.de

Turk, M., & Pentland, A. (1991). *Eigenface for Recognition, 3*(1), 71-86.

Appendix

Appendix 1.0

This section contains the entire study database collected along the various environmental constraints. The six constraints (Angry, Disgust, Fear, Happy, Sad and Surprise) were used for testing the algorithms whereas the neutral expressions were trained and knowledge captured in memory for recognition.

Appendix 1.0a: Cohn Kanade Face Expression database with the various study constraints.

Appendix 1.0a: Cohn Kanade Face Expression database with the various study constraints.

Appendix 1.0a: Cohn Kanade Face Expression database with the various study constraints.

Appendix 1.0b: Ghanaian Face Expression database with the various study constraints.

Appendix 1.0c: JAFFE database with the various study constraints.

Appendix 2.0

This section contains the multivariate data from each of the study algorithms. Here the data been presented are the absolute deviations from the variates means.

Appendix 2.0a: Multivariate data from Algorithm 1

Indiv.	Angry	Disgust	Fear	Happy	Sad	Surprise
1	3517.1	3312.4	3138.9	382.56	3200.9	3416.7
2	3648.5	3188.9	1369.8	1270	977.53	1200.8
3	1066.7	2292.3	944.91	1754.1	153.65	503.84
4	1887.2	1565.2	1968.2	650.66	1592.4	3163.7
5	1310	1786.4	714.62	1555.7	1041.4	19.272
6	3344	2233.9	396.92	3125.5	1639.3	457.94
7	2026	3846	2079.9	2066.4	3774.7	1257.6
8	3938.2	1352.1	287.93	603.52	2255.8	2297.1
9	4511.8	2741.5	2001.4	612.89	343.57	555.98
10	3911.8	2740.3	431.02	1892.2	1288.2	3438.5
11	4061.4	3992.9	2417.5	3841.7	3093.6	3190.7
12	1007.3	311.05	1354.2	2343.1	60.553	4302.2
13	641.44	1039.4	1502.3	2711.9	20.156	439.59
14	4498.9	213.6	1725.3	381.99	3122.7	389.05
15	1291.7	649.23	2096	2441.9	2137.4	1073.8
16	604.4	662.43	2188.9	450.77	38.288	150.02
17	1592.4	726.54	972.45	3186.6	2050.1	1685.6
18	2443.2	2156.4	968.9	2.4806	1099.1	1322.6
19	2367.3	2632.2	1758.5	3362.8	602.98	2763.2
20	4202.9	1733.3	2035.5	802.91	3368.7	2945.4

Appendix 2.0a: Multivariate data from Algorithm 1

Indiv.	Angry	Disgust	Fear	Happy	Sad	Surprise
21	2649.6	2807.1	2181.8	388.48	2203	1093.8
22	435.71	1929.4	115.94	1088.9	1845.6	5610.8
23	974.95	1838.1	21.981	2655.7	3448.7	713.35
24	2468.7	2045.5	3403	961.96	1571.8	2048.3
25	3914.6	2149	2617.9	2952	1991.8	2543.7
26	1960.6	832.35	986.1	1915.1	716.68	1099.8
27	8971.9	2449.4	249.59	878.3	3907	517.22
28	13519	10854	421.22	14982	1501.9	6786.5
29	4172.6	2069.4	2502.5	1358.1	900.82	1927.7
30	16692	703.38	4804.2	2626.1	751.38	2277
31	497.09	3554.5	537.24	1834.1	3030.6	978.56
32	577.03	1869.1	6.0795	1428.3	342.6	487.2
33	1877	2814.1	3018.8	2190.9	430.11	2250.2
34	1876.8	3680.2	3528.1	2197.5	1940.2	2460.5
35	540.55	977.6	3713.8	1454.1	4722.7	3596.4
36	598.35	1004.1	247.22	1411.1	662	1216.5
37	2560.7	1592.4	365.71	657.07	2981.1	1623.5
38	4335.3	6486.2	2552.9	1564.1	4242.9	680.9
39	3086.1	9049.1	1717.6	3712.9	3998.4	1340.5
40	490.52	239.92	1581.3	296.72	2493.8	1591.7
41	312.47	1362.2	2670.1	4459.2	2594.1	180.96
42	7593.3	3786.5	2211.3	6672.8	7580.7	4450.4

Appendix 2.0b: Multivariate data from Algorithm 2

Indiv.	Angry	Disgust	Fear	Happy	Sad	Surprise
1	3929.3	1918	2907.8	984.55	1742	2248.9
2	1637	4053.6	921.14	152.62	122.54	114.62
3	2989.6	945.27	1693.9	84.026	3856.2	3642.9
4	1741.7	1329	2337.7	272.14	2404.7	4884.5
5	5219	7353.3	5542.9	7505.9	6778.4	4785
6	4900.5	4418.6	1258.3	7069.3	2382.7	1736
7	2098.5	6856.2	64.71	749.39	4640.6	1429.5
8	8407.7	7262.5	1341.1	2030	8003.2	7156
9	3915.7	5193.4	5625.6	6103.2	5826	4427
10	3906.4	1869	2432.6	42.758	254.16	1532.7
11	7144.2	7174.7	7001.7	7722.8	7771.1	6110.5
12	177.04	562.65	1252.1	1229.1	884.5	3956.1
13	4545.3	6205.1	3990.9	1141.5	8371.8	5793.5
14	6672.5	5363.1	2682	3198.5	5694.5	2677.7
15	1094.5	3998.4	3502.7	2484.9	1843	2839.9
16	4918.5	5932.7	1625.5	3949.6	4863	2854.8
17	2627.6	4213.1	2539.8	751.27	6783	352.22
18	12174	14077	3140.7	8239	10686	17268
19	3018.2	1995.5	2996.9	1304.1	4832	2564.1
20	1505.1	2684.6	977.25	4505.3	1772.7	1420.1

Appendix 2.0b: Multivariate data from Algorithm 2

Indiv.	Angry	Disgust	Fear	Happy	Sad	Surprise
21	6934.2	10366	8812.9	3080.2	5346.2	907.24
22	5300.4	11536	13774	8873.7	9510.6	2815.2
23	3083.1	10613	5415.7	11820	14835	11604
24	3338.9	3070.3	3024.4	1157.2	1415.9	574.77
25	7904.7	5899.9	7441.7	7210	7234.5	5957
26	5393.3	2711	3402.5	6448.5	2097.5	2860.9
27	5950.3	6829.8	1126.5	951.82	5932.9	5839.2
28	1158	881.27	1109.7	5.1759	1951.9	1178.9
29	9067.7	8645.5	9289.1	9909.8	8146.8	6254
30	330.9	6931.3	8995.5	13328	1857.6	2752.8
31	3205.9	3275.2	8366.8	7883.2	7951.6	4761.1
32	3818	4746.1	8554.9	3525.5	9534	4869.3
33	3071.8	6427.1	3628.7	6806.8	7390.1	2400.4
34	3422.8	4109.7	2993.2	3249.8	3521.3	1940.9
35	2826.6	703.15	20.182	2048	373.76	864.58
36	1199.3	719.29	6833.2	9573.9	6802.5	9171.2
37	6985.2	7310.3	4305.4	7072.3	7992.7	5140.8
38	737.83	119.12	4402.5	931.8	1390.3	3810.9
39	4518.2	6239.7	3104.1	4036.8	4293.3	2421
40	304.12	25.3	1876.7	4268.7	6646.2	6396.3
41	4235.9	4675.4	5187.6	2814.8	5867.4	2581.4
42	4217	2560.2	3816.8	4485.8	4647.4	2584.2

International Journal of Statistics and Probability; Vol. 4, No. 4; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

The Exponentiated Burr XII Poisson Distribution with Application to Lifetime Data

Ronaldo V. da Silva[1], Frank Gomes-Silva[2], Manoel Wallace A. Ramos[3] & Gauss M. Cordeiro[4]

[1] Recife Military School, Recife, PE, Brazil

[2] Department of Statistics and Informatics, Federal Rural University of Pernambuco, Recife, PE, Brazil

[3] Federal Institute of Paraíba, João Pessoa, PB, Brazil

[4] Department of Statistics, Federal University of Pernambuco, Recife, PE, Brazil

Correspondence: Frank S. Gomes da Silva, Department of Statistics and Informatics, Federal Rural University of Pernambuco, Rua Dom Manoel de Medeiros, s/n, Campus Dois Irmãos, 52171-900, Recife, Pernambuco, Brazil. Tel: 81-3320-6491. E-mail: franksinatrags@gmail.com

Received: August 13, 2015 Accepted: August 27, 2015 Online Published: October 9, 2015

doi:10.5539/ijsp.v4n4p112 URL: http://dx.doi.org/10.5539/ijsp.v4n4p112

Abstract

A five-parameter distribution, called the exponentiated Burr XII Poisson distribution, is defined and studied. The model has as special sub-models some important lifetime distributions discussed in the literature, such as the logistic, log-logistic, Weibull, Burr XII and exponentiated Burr XII distributions, among several others. We derive the ordinary and incomplete moments, generating and quantile functions, Bonferroni and Lorenz curves, mean deviations, reliability and two types of entropy. The order statistics and their moments are investigated. The method of maximum likelihood is proposed for estimating the model parameters. We obtain the observed information matrix. An application to a real data set demonstrates that the new distribution can provide a better fit than other classical lifetime models. We hope that this generalization may attract wider applications in reliability, biology and survival analysis.

Keywords: Beta distribution, Burr XII distribution, Maximum likelihood, Observed information matrix, Weibull distribution

1. Introduction

The statistics literature has numerous distributions for modeling lifetime data. But many if not most of these distributions lack motivation from a lifetime context. For example, there is not apparent physical motivation for the gamma distribution. It only has a more general mathematical form than the exponential distribution with one additional parameter, so it has nicer properties and provides better fits. The same arguments apply to the BXII distribution, among others.

Zimmer *et al.* (1998) introduced the three parameter Burr XII (BXII) distribution with cumulative distribution function (cdf) and probability density function (pdf) (for $x > 0$) given by

$$G(x; s, k, c) = 1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k} \tag{1}$$

and

$$g(x; s, k, c) = c\, k\, s^{-c}\, x^{c-1} \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1}, \tag{2}$$

respectively, where $k > 0$ and $c > 0$ are shape parameters and $s > 0$ is a scale parameter. If $c > 1$, the density function is unimodal with mode at $x = s\,[(c-1)/(ck+1)]^{1/c}$ and is L-shaped if $c = 1$. If $q < c\,k$, the qth moment about zero is $\mu'_q = s^q\, k\, B(k - q\,c^{-1}, 1 + q\,c^{-1})$, where $B(p, q) = \Gamma(p)\,\Gamma(q)/\,\Gamma(p+q)$ and $\Gamma(p) = \int_0^\infty x^{p-1}\, \mathrm{e}^{-x}\, \mathrm{d}x$ is the gamma function.

The BXII distribution, having as sub-models the logistic and Weibull distributions, is a very popular distribution for modeling lifetime data and phenomenon with monotone failure rates. When modeling monotone hazard rates, the Weibull model may be an initial choice because of its negatively and positively skewed density shapes. Nevertheless, it does not furnish a reasonable parametric fit for non-monotone failure rates such as the bathtub shaped and unimodal failure rates that are common in reliability and biological studies.

Several other authors including El-Bassiouny and Abdo (2010), Jayakumar and Mathew (2008), Brito *et al.* (2014) and Ramos *et al.* (2015) proposed and developed the structural properties of various generalized Burr XII distributions.

The cdf and the reliability function of the three-parameter BXII distribution can be expressed in closed-form, thus simplifying the computation of the percentiles and the likelihood function for censored data. This distribution has algebraic tails that are effective for modeling failures occurring with lesser frequency than with those models based on exponential tails. Hence, it represents an adequate distribution for modeling failure time data (Zimmer *et al.*, 1998). Shao (2004) discussed maximum likelihood estimation of its parameters and Shao *et al.* (2004) studied models for extremes based on the BXII distribution with application to flood frequency analysis. According to Soliman (2005), this model generalizes a large number of distributions. Its versatility and flexibility turns it quite attractive as a tentative model for lifetime data.

For an arbitrary baseline cdf $G(x)$, a random variable is said to have the *exponentiated-G* ("Exp-G" for short) distribution with parameter $a > 0$, say $X \sim \text{Exp-}G(a)$, if its pdf and cdf are $H_a(x) = G^a(x)$ and $h_a(x) = aG^{a-1}(x)\, g(x)$, respectively. Thus, the cdf and pdf of the *exponentiated Burr XII* (Exp-BXII) distribution is given by

$$G_\alpha(x; s, k, c) = \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^\alpha \tag{3}$$

and

$$g_\alpha(x; s, k, c) = k\, c\, s^{-c}\, \alpha\, x^{c-1} \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1} \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^{\alpha-1}, \tag{4}$$

respectively.

We provide four motivations for the proposed lifetime model called the *exponentiated BXII Poisson* (Exp-BXIIP) distribution. The first is based on failures of a system. Suppose that a system has N serial sub-systems functioning independently at a give time, where N is a truncated Poisson random variable with probability mass function (pmf)

$$Pr(N = n) = \frac{1}{(e^\lambda - 1)} \frac{\lambda^n}{n!} \tag{5}$$

for $n = 1, 2, \ldots$ Let X denote the time of failure of the first out of the N functioning systems defined by the independent random variables $Y_1 \sim \text{Exp-BXII}(\alpha)$, $\ldots$, $Y_N \sim \text{Exp-BXII}(\alpha)$ given by the cdf (3). Then, $X = min(Y_1, \ldots, Y_N)$. So, the conditional cdf of X (for $x > 0$) given N is

$$\begin{aligned} F(x|N) &= 1 - Pr(X > x|N) = 1 - Pr(Y_1 > x, \ldots, Y_N > x) \\ &= 1 - Pr^N(Y_1 > x) = 1 - [1 - Pr(Y_1 \le x)]^N \\ &= 1 - \left\{1 - \left[1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right]^\alpha\right\}^N, \end{aligned}$$

where $s, k, c, \alpha, \lambda > 0$. Hence, the unconditional cdf of X is

$$\begin{aligned} F(x) &= \frac{1}{(e^\lambda - 1)} \sum_{n=1}^{\infty} \left\{1 - \left[1 - \left(1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right)^\alpha\right]^n\right\} \frac{\lambda^n}{n!} \\ &= \frac{1}{(1 - e^{-\lambda})} \left\{1 - \exp\left\{-\lambda \left[1 - \left(1 + \left(\frac{x}{s}\right)^c\right)^{-k}\right]^\alpha\right\}\right\}. \end{aligned} \tag{6}$$

Then,

$$F(x) = \frac{1}{(1 - e^{-\lambda})} \left\{1 - \exp\left[-\lambda\, G(x)^\alpha\right]\right\}, \tag{7}$$

where $G(x) = G(x; s, k, c)$ is given by (1). We refer to the distribution (6) as the Exp-BXIIP distribution. Proving a new lifetime distribution is always precious for statisticians. The fact that the new model generalizes existing commonly used distributions is also a positive point.

The survival function associated with X becomes

$$S(x) = 1 - F(x) = \frac{1}{(1 - \mathrm{e}^{-\lambda})} \left\{ \exp\left\{ -\lambda \left[1 - \left(1 + \left(\frac{x}{s} \right)^{c} \right)^{-k} \right]^{\alpha} \right\} - \exp(-\lambda) \right\}. \tag{8}$$

The probability density function (pdf) corresponding to (6) is given by

$$\begin{aligned} f(x; s, k, c, \alpha, \lambda) &= \frac{cks^{-c}\,\alpha\lambda}{1 - \mathrm{e}^{-\lambda}}\, x^{c-1} \left[1 + \left(\frac{x}{s} \right)^{c} \right]^{-k-1} \left\{ 1 - \left[1 + \left(\frac{x}{s} \right)^{c} \right]^{-k} \right\}^{\alpha-1} \\ &\times \exp\left\{ -\lambda \left[1 - \left(1 + \left(\frac{x}{s} \right)^{c} \right)^{-k} \right]^{\alpha} \right\}. \end{aligned} \tag{9}$$

Hereafter, a random variable X with density function (9) is denoted by $X \sim$ Exp-BXIIP$(s, k, c, \alpha, \lambda)$. Plots of the density function of X for selected parameter values are displayed in Figure 1.

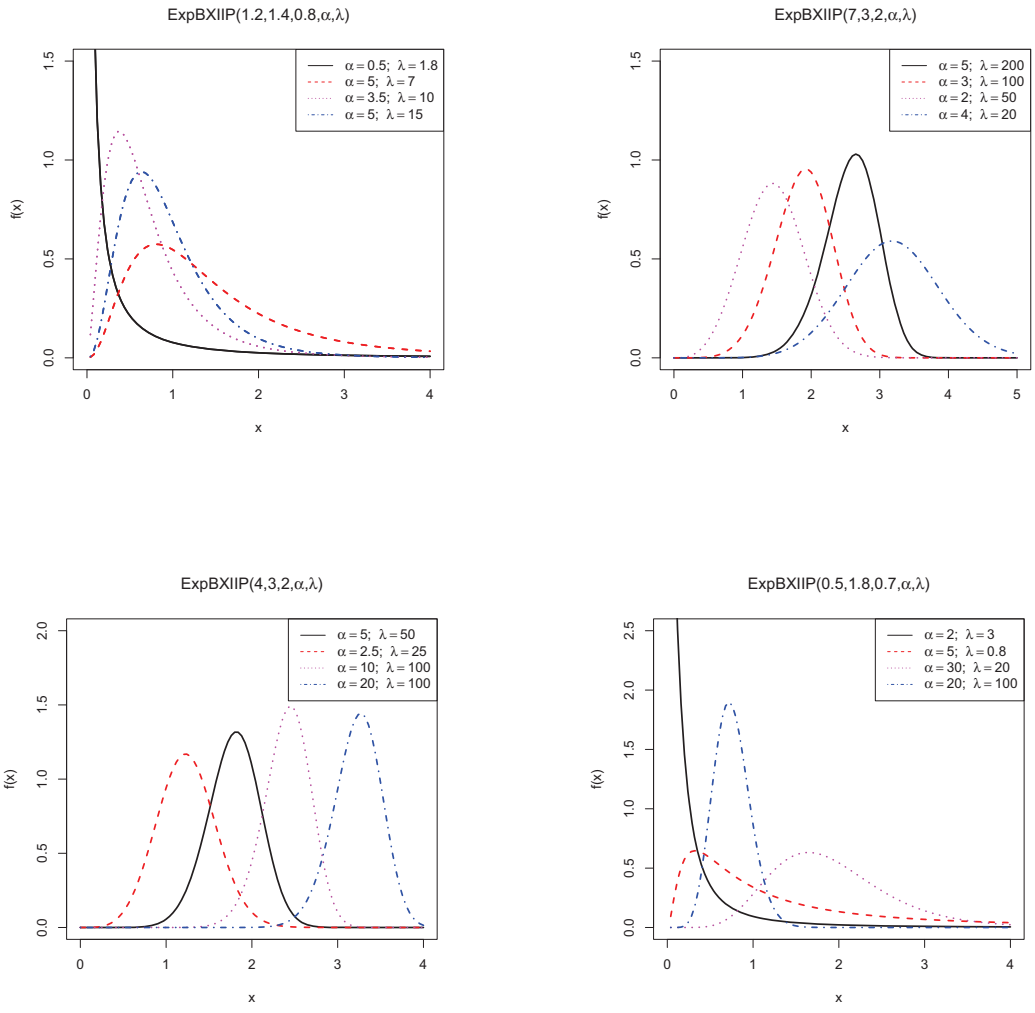


Figure 1. Plots for the Exp-BXIIP density for some parameter values.

The Exp-BXIIP hazard rate function (hrf) is given by

$$\begin{aligned}\tau(x;s,k,c,\alpha,\lambda) &= \frac{c\,k\,s^{-c}\,\alpha\lambda\,x^{c-1}\,[1+(\frac{x}{s})^c]^{-k-1}\,\{1-[1+(\frac{x}{s})^c]^{-k}\}^{\alpha-1}}{\exp\{-\lambda[1-(1+(\frac{x}{s})^c)^{-k}]^{\alpha}\}-\exp(-\lambda)} \\ &\times \exp\{-\lambda[1-(1+(\frac{x}{s})^c)^{-k}]^{\alpha}\}. \end{aligned} \tag{10}$$

Plots of the hazard rate functions for selected parameter values are displayed in Figure 2.

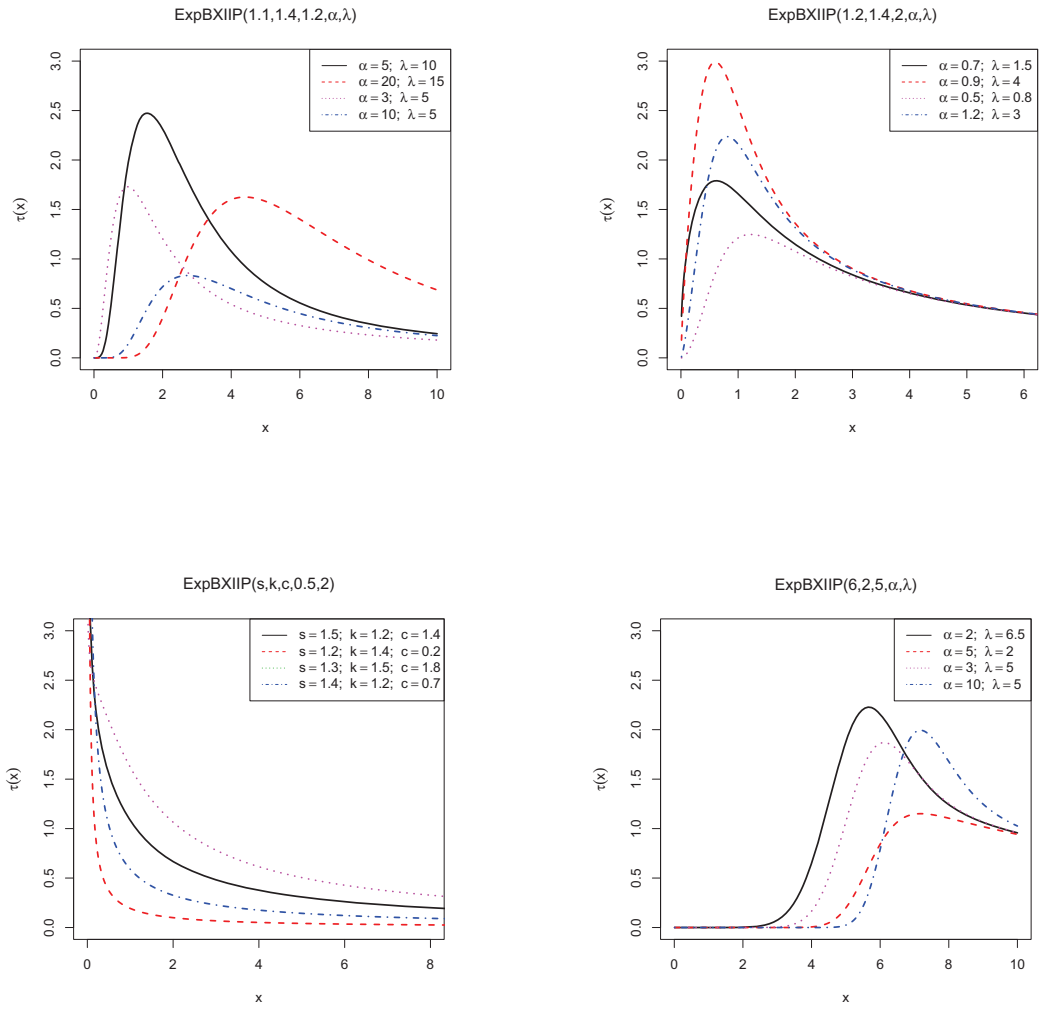


Figure 2. The Exp-BXIIP hrf for some parameter values.

For a second motivation suppose that an *ith* system is made of α parallel components, so that the system will fail if all of the components fail. Assume that the failure times of the components for the *ith* system, say $Z_{i,1}, Z_{i,2}, \ldots, Z_{i,\alpha}$, are independent and identically BXII random variables with parameters s, k, c. Let Y_i denote the failure time of the *ith* system and that there is an unknown number N of independent systems. The cdf of the failure time X of the first system out of the N functioning system is given by (6).

For the third motivation, we assume that N is the unknown number of carcinogenic cells for an individual left active after the initial treatment has pmf (5) and that Y_i is the time spent for the ith carcinogenic cell to produce a detectable

cancer mass. Assuming that $Y_1, \ldots, Y_N$ is a sequence of iid Exp-BXII random variables independent of N, the time to relapse of cancer of a susceptible individual can be modeled by the Exp-BXIIP family of distributions.

Finally, the fourth motivation considers that the failure of a device occurs due to the presence of an unknown number N of initial defects of the same kind, which can be identifiable only after causing failure and are repaired perfectly. Define by Y_i the time to the failure of the device due to the ith defect, for $i \geq 1$. If we assume that the Y_i's are iid Exp-BXII random variables independent of N having pmf (5), then the time to the first failure is appropriately modeled by the Exp-BXIIP distribution. For reliability studies, the Exp-BXIIP models can arise in series and parallel systems with identical components, which appear in many industrial applications and biological organisms. These points indicate that the new family of distributions is well-motivated for industrial applications and biological studies.

In this paper, we study some mathematical properties of the Exp-BXIIP model and illustrate its potentiality. In Section 2, we demonstrate that the cdf and pdf of X can be expressed as a mixture of Exp-BXII densities. Explicit expressions for the ordinary and incomplete moments are derived in Section 3. Generating and quantile functions are derived in Section 4 and 5, respectively. In Section 6, mean deviations and reliability are derived. In Section 7, we investigate the order statistics and some of their structural properties. Rényi and Shannon entropies are derived in Section 8. Maximum likelihood estimation of the model parameters is performed and the observed information matrix is determined in Section 9. In Section 10, we provide an application of the Exp-BXIIP to a real data set. Finally, Section 11 ends with some concluding remarks.

2. Useful Expansions

Using the Taylor series

$$1 - \mathrm{e}^{-\mathrm{z}} = \sum_{\mathrm{k}=1}^{\infty} \frac{(-1)^{\mathrm{k}+1}\mathrm{z}^{\mathrm{k}}}{\mathrm{k}!},$$

equation (6) can be expressed as

$$F(x) = \frac{1}{(1-\mathrm{e}^{-\lambda})} \sum_{j=0}^{\infty} \frac{(-1)^j \lambda^{j+1}}{(j+1)!} \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^{(j+1)\alpha}, \tag{11}$$

and then

$$F(x) = \sum_{j=0}^{\infty} \omega_j \, H_{(j+1)\alpha}(x; s, k, c), \tag{12}$$

where $\omega_j = \frac{(-1)^j \lambda^{j+1}}{(j+1)!\,(1-\mathrm{e}^{-\lambda})}$ and $H_\alpha(x; s, k, c) = G^\alpha(x; s, k, c)$ is the Exp-BXII cdf. Clearly, $\sum_{j=1}^{\infty} w_j = 1$.

By differentiating (12), we can write

$$f(x) = \sum_{j=0}^{\infty} \omega_j \; h_{(j+1)\alpha}(x; s, k, c), \tag{13}$$

where $h_{(j+1)\alpha}(x; s, k, c)$ denotes the Exp-BXII fdp with parameters s, k, c and power parameter $(j+1)\alpha$. Equation (13) reveals that the Exp-BXIIP density function is a mixture of Exp-BXII densities.

3. Properties

Some of the most important features and characteristics of a distribution can be studied through moments (e.g., tendency, dispersion, skewness and kurtosis).

Theorem 1 *If $X \sim$ Exp-BXIIP$(s, k, c, \alpha, \lambda)$, we have the following approximations:*

1.1 *For $\alpha > 0$ and $\lambda > 0$ real non-integers, we have the mixture representation*

$$f(x) = \sum_{r=0}^{\infty} v_r \, g(x; s, k(r+1), c), \tag{14}$$

where $g(x; s, k(r+1), c)$ denotes the BXII density function with scale parameter s and shape parameters c and $k(r+1)$, and the coefficients are given by

$$v_r = \frac{\alpha\,\lambda}{(r+1)!\,(1-\mathrm{e}^{-\lambda})} \sum_{j=0}^{\infty} \frac{(-1)^{j+r}\,\lambda^j\,\Gamma[(j+1)\alpha]}{j!\,\Gamma[(j+1)\alpha - r]}. \tag{15}$$

Clearly, $\sum_{r=0}^{\infty} v_r = 1$. *Equation (14) reveals that the Exp-BXIIP density function is an infinite linear combination of BXII density functions. So, some structural properties of the Exp-BXIIP distribution can be obtained from those of the BXII distribution.*

1.2 *For* $\alpha > 0$ *and* $\lambda > 0$ *real non-integers, we obtain*

$$F(x) = \sum_{r=0}^{\infty} v_r \, G(x; s, k(r+1), c). \tag{16}$$

1.3 *If* $n < kc$, *the nth ordinary moment of the Exp-BXIIP distribution is given by*

$$\mu'_n = E(X^n) = k \, s^n \sum_{r=0}^{\infty} v_r \, B\Big[k(r+1) - \frac{n}{c}, \frac{n}{c} + 1\Big]. \tag{17}$$

Proof 1.1.

First, if $z \in \mathbb{R}$, we have the power series

$$e^{-z} = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} z^j. \tag{18}$$

Second, if $|z| < 1$ and b is a nonnegative integer, the power series holds

$$(1-z)^{b-1} = \sum_{j=0}^{\infty} \frac{(-1)^j \, \Gamma(b)}{\Gamma(b-j) \, j!} \, z^j. \tag{19}$$

Using (18), the Exp-BXIIP density function (9) can be expressed as

$$f(x) = \frac{cks^{-c} \, \alpha\lambda}{(1-\mathrm{e}^{-\lambda})} \, x^{c-1} \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k-1} \sum_{j=0}^{\infty} \Big[\frac{(-1)^j \, \lambda^j}{j!} \tag{20}$$

$$\times \left\{1 - \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k}\right\}^{(j+1)\alpha-1}\Big]. \tag{21}$$

Further, using (19), we obtain

$$f(x) = \frac{ck(r+1)s^{-c} \, \alpha\lambda}{(1-\mathrm{e}^{-\lambda})} \, x^{c-1} \sum_{j,r=0}^{\infty} \Big\{\frac{(-1)^{j+r} \, \lambda^j \, \Gamma[(j+1)\alpha]}{j! \, (r+1)! \, \Gamma[(j+1)\alpha - r]} \tag{22}$$

$$\times \left[1 + \left(\frac{x}{s}\right)^c\right]^{-k(r+1)-1}\Big\}.$$

Finally, we have

$$f(x) = \sum_{r=0}^{\infty} v_r \, g(x; s, k(r+1), c),$$

where v_r is given by (15) and $g(x; s, k(r+1), c)$ was defined before.

Proof 1.2.

Using Theorem 1.1 we obtain (16) by simple integration.

Proof 1.3.

The nth moment of X comes from Theorem 1.1

$$\mu'_n = \sum_{r=0}^{\infty} v_r \, E(Y_{r+1}), \tag{23}$$

where $Y_{r+1} \sim \text{BXII}(s, k(r+1), c)$. Using a result in Zimmer *et al.* (1998), we obtain for $n < kc$

$$\mu'_n = k\, s^n \sum_{r=0}^{\infty} v_r\, B[k(r+1) - nc^{-1}, nc^{-1} + 1].$$

The central moments (μ_s) and cumulants (κ_s) of X can be determined from (17) as

$$\mu_s = \sum_{k=0}^{p} (-1)^k \binom{s}{k} \mu_1'^s\, \mu'_{s-k} \qquad \text{and} \qquad \kappa_s = \mu'_s - \sum_{k=1}^{s-1} \binom{s-1}{k-1} \kappa_k\, \mu'_{s-k},$$

respectively, where $\kappa_1 = \mu'_1$.

For lifetime models, it is usually of interest to compute the nth incomplete moment of X defined by $m_n(y) = \int_0^y x^n\, f(x)dx$. The quantity $m_n(y)$ can be calculated from (14) as

$$m_n(y) = k\, c \sum_{r=0}^{\infty} (r+1)\, v_r \int_0^y x^{n-1} \left(\frac{x}{s}\right)^c \left[1 + \left(\frac{x}{s}\right)^{-k(r+1)-1}\right] dx.$$

Setting $t = \left[1 + \left(\frac{x}{s}\right)^c\right]^{-1}$, we can write

$$m_n(y) = k\, s^n \sum_{r=0}^{\infty} (r+1)\, v_r \int_0^{\frac{s^c}{s^c+y^c}} t^{k(r+1)-\frac{n}{c}-1} (1-t)^{\frac{n}{c}} dt$$

and then for $n < kc$

$$m_n(y) = k\, s^n \sum_{r=0}^{\infty} (r+1)\, v_r\; B_{\frac{s^c}{s^c+y^c}}\Big(k(r+1) - n\, c^{-1}, 1 + n\, c^{-1}\Big), \tag{24}$$

where $B_z(a,b) = \int_0^z t^{a-1}\,(1-t)^{b-1} dt$ is the incomplete beta function.

4. Moment Generating Function

An explicit expression for $M(t)$ can be obtained from equation (14) as an infinite weighted sum

$$M(t) = \sum_{r=0}^{\infty} v_r\, M_{r+1}(t), \tag{25}$$

where $M_{r+1}(t)$ is the moment generating function (mgf) of Y_{r+1} and v_r is defined by (15). We provide a simple representation for the mgf $M_{\text{BXII}}(t)$ of the BXII(s,k,c) distribution. We can write for $t < 0$

$$M_{\text{BXII}}(t) = ck \int_0^{\infty} \mathrm{e}^{\mathrm{y}\,t}\, \mathrm{y}^{\mathrm{c}-1} (1 + \mathrm{y}^{\mathrm{c}})^{-(\mathrm{k}+1)} \mathrm{dy}.$$

Now, we use the Meijer G-function defined by

$$G_{p,q}^{m,n}\left(x \,\middle|\, \begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix}\right) = \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^{m} \Gamma\left(b_j + t\right) \prod_{j=1}^{n} \Gamma\left(1 - a_j - t\right)}{\prod_{j=n+1}^{p} \Gamma\left(a_j + t\right) \prod_{j=m+1}^{p} \Gamma\left(1 - b_j - t\right)} x^{-t} dt,$$

where $\mathrm{i} = \sqrt{-1}$ is the complex unit and L denotes an integration path; see Section 9.3 in Gradshteyn and Ryzhik (2000) for a description of this path. The Meijer G-function contains as particular cases many integrals with elementary and special functions (Prudnikov *et al.*, 1986).

We now assume that $c = m/k$, where m and k are positive integers. This condition is not restrictive since every positive real number can be approximated by a rational number. Using the integral (38) given in Appendix A, we conclude for $t < 0$ that

$$M_{\text{BXII}}(t) = mI\left(-st, \frac{m}{k} - 1, \frac{m}{k}, -k - 1\right). \tag{26}$$

Now, from equation (25), the mgf of the Exp-BXIIP$(s, k, c, \alpha, \lambda)$ distribution (for $t < 0$) follows as

$$M(t) = m \sum_{r=0}^{\infty} v_r \, I\left(-st, \frac{m}{k(r+1)} - 1, \frac{m}{k(r+1)}, -k(r+1) - 1\right). \tag{27}$$

Equation (27) is the main result of this section. For the special cases $c = 1$ and $c = 2$, we can obtain simple expressions for $M_{\mathrm{BXII}}(t)$ and, consequently, for $M(t)$ using equations (1) (on page 16) and (2) (on page 20) of the book by Prudnikov *et al.* (1992). For $c = 1$ and $t < 0$, we have

$$M_{\mathrm{BXII}}(t) = k(-st)^k \, \mathrm{e}^{-\mathrm{st}} \, \Gamma(-\mathrm{k}, -\mathrm{st}),$$

where $\Gamma(v, x) = \int_x^{\infty} t^{v-1} \mathrm{e}^{-\mathrm{st}} \mathrm{dt}$ is the complementary incomplete gamma function. For $c = 2$ and $t < 0$, we obtain

$$\begin{aligned} M_{\mathrm{BXII}}(t) &= {}_1F_2\Big(1; \frac{1}{2}; 1-k; \frac{s^2t^2}{4}\Big) + \frac{st}{2} B\Big(2, k - \frac{1}{2}\Big) \, {}_1F_2\Big(1; \frac{3}{2}; k + \frac{7}{2}; \frac{-s^2t^2}{4}\Big) \\ &+ \frac{\Gamma(-2k)}{(-st)^{-2k}}, \end{aligned}$$

where

$${}_1F_2(a, b; c; x) = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k (c)_k} \frac{x^k}{k!}$$

is a generalized hypergeometric function and $(a)_k = a(a+1)\ldots(a+k-1)$ denotes the ascending factorial.

5. Quantile Function

The Exp-BXIIP quantile function, say $x = Q(u)$, can be obtained by inverting (6). We have

$$x = Q(u) = F^{-1}(u) = s\left\{\left[1 - \left\{-\lambda^{-1} \log\left[1 - u(1 - \mathrm{e}^{-\lambda})\right]\right\}^{\frac{1}{\alpha}}\right]^{\frac{-1}{k}} - 1\right\}^{\frac{1}{c}}. \tag{28}$$

The shortcomings of the classical kurtosis measure are well-known. For example, the moments of X in (9) are valid only for $n < kc$. There are many heavy-tailed distributions for which this quantile is infinite. So, it becomes uniformative precisely when it needs to be. Indeed, our motivation to use quantile-based measures stemmed from the non-existence of classical kurtosis for many generalized distributions. The Bowley skewness (see Kenney and Keeping, 1962) is based on quartiles

$$B = \frac{Q(3/4) - 2Q(1/2) + Q(1/4)}{Q(3/4) - Q(1/4)}$$

whereas the Moors kurtosis (see Moors, 1998) is based on octiles

$$M = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(6/8) - Q(2/8)},$$

where $Q(\cdot)$ denotes the Exp-BXIIP quantile function given by (28). Plots of the B and M functions for selected parameter values are displayed in Figure 3.

6. Other Measures

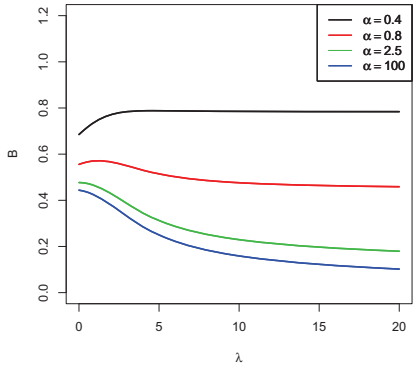
In this section, we calculate the following measures: means deviations, Bonferroni and Lorenz curves and the reliability of the Exp-BXIIP distribution.

6.1 Mean Deviations

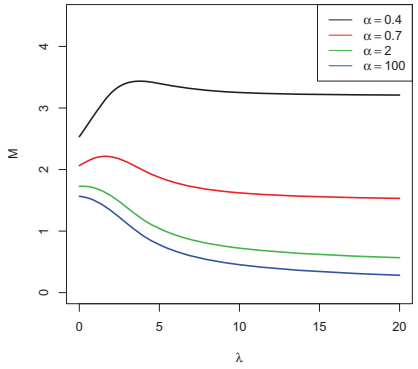
Here, we determine the mean deviations and Bonferroni and Lorenz curves of X. The amount of scatter in a population is evidently measured to some extent by the totality of deviations from the mean and median. These are known as the mean deviation about the mean and the mean deviation about the median – defined by

$$\delta_1(X) = 2\mu_1' \, F(\mu_1') - 2m_1(\mu_1') \qquad \text{and} \qquad \delta_2(X) = \mu_1' - 2m_1(M),$$

respectively, where $\mu_1' = E(X)$, $F(\mu_1')$ is obtained from (6), the median M of X is calculated from the quantile function (28) by $M = Q(1/2)$ and $m_1(q) = \int_0^q x \, f(x) dx$ is the incomplete mean of X given by (24) with $n = 1$.



(a) The B function for $s = 1.2$, $k = 1.4$ and $c = 0.8$.



(b) The M function for $s = 1.2$, $k = 1.4$ and $c = 0.8$.

Figure 3. Plots of the B and M functions for some parameter values

Setting $u = y^c$, we can write from equation (14)

$$m_1(q) = k\, s \sum_{r=0}^{\infty} (r+1)\, v_r \int_{(q/s)^c}^{\infty} u^{1/c}\, (1+u)^{-k(r+1)-1} du,$$

where the integral can be calculated using Maple as

$$m_1(q) = k\, s \sum_{r=0}^{\infty} (r+1)\, v_r\; J\left(\left[\frac{q}{s}\right]^c, \frac{1}{c}, k(r+1)+1\right). \tag{29}$$

Here,

$$\begin{aligned} J(q,r,k) &= \int_q^{\infty} u^r\, (1+u)^{-k} du \\ &= -\left[\frac{{}_2F_1[(k, r+1); (2+r); -q] q^{r+1}}{(r+1)} + \frac{\pi\, \Gamma(k-r-1)\, \csc(\pi r)}{\Gamma(k)\Gamma(-r)}\right], \end{aligned}$$

where $\csc(\cdot)$ is the cosecant function and ${}_2F_1$ is the hypergeometric function defined by

$${}_2F_1(a, b; c; x) = \sum_{k=0}^{\infty} \frac{(a)_k\, (b)_k}{(c)_k}\, \frac{x^k}{k!}.$$

Equation (29) is the main result of this section from which $\delta_1(X)$ and $\delta_2(X)$ are immediately determined. The mean deviations can be used to plot Lorenz and Bonferroni curves in fields like economics, reliability, demography, insurance and medicine. For a given probability π, they are defined by $L(\pi) = m_1(q)/\mu_1'$ and $B(\pi) = m_1(q)/(\pi\, \mu_1')$, respectively, where $q = Q(\pi)$ comes directly from (28).

6.2 Reliability

In reliability, the stress-strength model describes the life of a component which has a random strength X_1 that is subjected to a random stress X_2. The component fails at the instant that the stress applied to it exceeds the strength, and the component will function satisfactorily whenever $X_1 > X_2$. Hence, $R = Pr(X_2 < X_1)$ is a measure of component reliability. It has many applications especially in engineering concepts, economics and physical science. We derive the reliability R when X_1 and X_2 have independent Exp-BXIIP$(s, k_1, c, \alpha_1, \lambda_1)$ and Exp-BXIIP$(s, k_2, c, \alpha_2, \lambda_2)$ distributions with identical scale parameter s and shape parameter c. The reliability is given by

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

The cdf of X_2 and density of X_1 are obtained from Theorem 1

$$F_2(x) = \sum_{r=0}^{\infty} v_r(\alpha_2, \lambda_2)\, G(x; s, k_2(r+1), c)$$

and

$$f_1(x) = \sum_{q=0}^{\infty} v_q(\alpha_1, \lambda_1)\, g(x; s, k_1(q+1), c).$$

Hence,

$$R = \frac{c\,k_1}{s^c} \sum_{r,q=0}^{\infty} (q+1)\, v_q(\alpha_1, \lambda_1)\, v_r(\alpha_2, \lambda_2)\, I(c, s, k_1, k_2, r, q),$$

where

$$\begin{aligned} I(c, s, k_1, k_2, r, q) &= \int_0^{\infty} \Big(x^{c-1} \Big[1 + \Big(\frac{x}{s}\Big)^c\Big]^{-k_1(q+1)-1} \\ &\times \Big\{1 - \Big[1 + \Big(\frac{x}{s}\Big)^c\Big]^{-k_2(r+1)}\Big\}\Big) dx. \end{aligned}$$

Setting $u = 1 + (x/s)^c$, we have

$$I(c, s, k_1, k_2, r, q) = \frac{ck_2(r+1) - 1}{ck_1(q+1) + ck_2(r+1)},$$

and then we obtain R.

7. Order Statistics

We now derive an explicit expression for the density of the ith order statistic $X_{i:n}$, say $f_{i:n}(x)$, in a random sample of size n from the Exp-BXIIP distribution. It is well-known that

$$f_{i:n}(x) = \frac{1}{B(i, n-i+1)} f(x) F(x)^{i-1} [1 - F(x)]^{n-i},$$

for $i = 1, \ldots, n$. Using the binomial expansion in the last equation, we readily obtain

$$f_{i:n}(x) = \sum_{l=0}^{n-i} \frac{(-1)^l \binom{n-i}{l} f(x)}{B(i, n-i+1)} F(x)^{i+l-1}. \tag{30}$$

We use the identity for k, v positive integer

$$\Big(\sum_{k=0}^{\infty} a_k\, x^k \Big)^v = \sum_{k=0}^{\infty} c_{v,k}\, x^k, \tag{31}$$

where $c_{v,0} = a_0^v$ and

$$c_{v,j} = \frac{1}{j\, a_0} \sum_{q=1}^{j} [(vq - j + q)\, a_q\, c_{v,j-q}]. \tag{32}$$

Then, we can write

$$\begin{aligned} F^v(x) &= \frac{1}{(1 - e^{-\lambda})^v} \Big\{1 - \Big[1 + \Big(\frac{x}{s}\Big)^c\Big]^{-k}\Big\}^{v\alpha} \Big\{ \sum_{j=0}^{\infty} \frac{(-1)^j\, \lambda^{j+1}}{(j+1)!} \\ &\times \Big\{1 - \Big[1 + \Big(\frac{x}{s}\Big)^c\Big]^{-k}\Big\}^{j\alpha} \Big\}^v \\ &= \frac{1}{(1 - e^{-\lambda})^v} \Big\{1 - \Big[1 + \Big(\frac{x}{s}\Big)^c\Big]^{-k}\Big\}^{v\alpha} \sum_{j=0}^{\infty} c_{v,j} \Big\{1 - \Big[1 + \Big(\frac{x}{s}\Big)^c\Big]^{-k}\Big\}^{j\alpha} \\ &= \frac{1}{(1 - e^{-\lambda})^v} \sum_{j=0}^{\infty} c_{v,j} \Big\{1 - \Big[1 + \Big(\frac{x}{s}\Big)^c\Big]^{-k}\Big\}^{(j+v)\alpha}, \end{aligned} \tag{33}$$

where $c_{v,j}$ is given in (32) and $a_j = \frac{(-1)^j \lambda^{j+1}}{(j+1)!}$.

Setting $v = i + l - 1$ and substituting (20) and (33) into equation (30), the density $f_{i:n}(x)$ can be expressed as

$$\begin{aligned} f_{i:n}(x) &= \frac{1}{(1-e^{-\lambda})^{i+l-1}} \sum_{l=0}^{n-i} \sum_{j=0}^{\infty} \Big[\frac{(-1)^l \binom{n-i}{l} c_{i+l-1,j}\, f(x)}{B(i, n-i+1)} \\ &\times \Big\{ 1 - \Big[1 + \Big(\frac{x}{s}\Big)^c \Big]^{-k} \Big\} \Big]^{(i+j+l-1)\alpha} \\ &= \frac{1}{(1-e^{-\lambda})^{i+l}} \sum_{l=0}^{n-i} \sum_{j,m=0}^{\infty} \frac{(-1)^{l+m} \binom{n-i}{l} \lambda^m\, c_{i+l-1,j}}{m!\, B(i, n-i+1)}\, c\, k\, s^{-c}\, \alpha\, \lambda\, x^{c-1} \\ &\times \Big[1 + \Big(\frac{x}{s}\Big)^c \Big]^{-k-1} \Big\{ 1 - \Big[1 + \Big(\frac{x}{s}\Big)^c \Big]^{-k} \Big\}^{(i+j+l+m)\alpha - 1} \\ &= \frac{1}{(1-e^{-\lambda})^{i+l}} \sum_{l=0}^{n-i} \sum_{j,m=0}^{\infty} \frac{(-1)^{l+m} \binom{n-i}{l} \lambda^{m+1}\, c_{i+l-1,j}}{(i+j+l+m)\, m!\, B(i, n-i+1)}\, c\, k\, s^{-c}\, \alpha \\ &\times (i+j+l+m)\, x^{c-1} \Big[1 + \Big(\frac{x}{s}\Big)^c \Big]^{-k-1} \Big\{ 1 - \Big[1 + \Big(\frac{x}{s}\Big)^c \Big]^{-k} \Big\}^{(i+j+l+m)\alpha - 1}. \end{aligned}$$

Thus, $f_{i:n}(x)$ can be written as

$$f_{i:n}(x) = \sum_{l=0}^{n-i} \sum_{j,m=0}^{\infty} \delta_{j,l,m}\, h_{(i+j+l+m)\alpha}(x), \tag{34}$$

where

$$\delta_{j,l,m} = \frac{(-1)^{l+m} \binom{n-i}{l} \lambda^{m+1}\, c_{i+l-1,j}}{(i+j+l+m)\, m!\, (1-\mathrm{e}^{-\lambda})^{\mathrm{i}+1}\, \mathrm{B(i, n-i+1)}},$$

and

$$\begin{aligned} h_{(i+j+l+m)\alpha}(x) &= c\, k\, s^{-c}\, (i+j+l+m)\, \alpha\, x^{c-1} \Big[1 + \Big(\frac{x}{s}\Big)^c \Big]^{-k-1} \\ &\times \Big\{ 1 - \Big[1 + \Big(\frac{x}{s}\Big)^c \Big]^{-k} \Big\}^{(i+j+l+m)\alpha - 1}. \end{aligned}$$

Thus, from equation (34), the tth ordinary moment of the Exp-BXIIP order statistics is

$$\mathrm{E}(X_{i:n}^t) = \sum_{l=0}^{n-i} \sum_{j,m=0}^{\infty} \delta_{j,l,m}\, \mathrm{E}(Y_{(i+j+l+m)\alpha}), \tag{35}$$

where $Y_{(i+j+l+m)\alpha} \sim$ Exp-BXII$(s, k, c, Y_{(i+j+l+m)\alpha})$. Clearly, $\mathrm{E}(X_{i,l}^t)$ can be calculated directly from equation (23) with the parameters of this Exp-BXII distribution.

An alternative expression to (35) can be derived using a result due to Barakat and Abdelkader (2004). We have

$$\mathrm{E}(X_{i:n}^t) = t \sum_{p=n-i+1}^{n} (-1)^{p-n+i-1} \binom{p-1}{n-i} \binom{n}{p} \int_0^{\infty} x^{t-1} S(x)^p dx, \tag{36}$$

where $S(x) = 1 - F(x)$ is the Exp-BXIIP survival function. Using the binomial expansion for $[1 - F(x)]^p$ in (36), the last integral becomes

$$L = \int_0^{\infty} x^{t-1} S(x)^p dx = \sum_{l=0}^{p} (-1)^l \binom{p}{l} \int_0^{\infty} x^{t-1} F(x)^l dx. \tag{37}$$

Substituting (33) into equation (37), we can rewrite L as

$$L = \sum_{l=0}^{p}\sum_{j=0}^{\infty}\frac{(-1)^l\binom{p}{l}c_{l,j}}{(1-\mathrm{e}^{-\lambda})^1}\int_0^{\infty}x^{t-1}\Big\{1-\Big[1+\Big(\frac{x}{s}\Big)^c\Big]^{-k}\Big\}^{(j+l)\alpha}dx,$$

where $c_{l,j}$ can be obtained from (32).

Using the power series expansion, we can write L as

$$L = \sum_{l=0}^{p}\sum_{j,m=0}^{\infty}\frac{(-1)^{l+m}\binom{p}{l}c_{l,j}\,\Gamma[(j+l)\,\alpha+1]}{m!\,(1-\mathrm{e}^{-\lambda})^1\,\Gamma[(\mathrm{j}+\mathrm{l})\,\alpha+1-\mathrm{m}]}\int_0^{\infty}x^{t-1}\Big[1+\Big(\frac{x}{s}\Big)^c\Big]^{-km}dx.$$

Setting $u = (x/s)^c$, we obtain

$$L = \sum_{l=0}^{p}\sum_{j,m=0}^{\infty}\frac{(-1)^{l+m}\binom{p}{l}s^t\,c_{l,j}\,\Gamma[(j+l)\,\alpha+1]}{c\,m!\,(1-\mathrm{e}^{-\lambda})^1\,\Gamma[(\mathrm{j}+\mathrm{l})\,\alpha+1-\mathrm{m}]}\Big[\frac{\Gamma(km-\frac{t}{c})\,\Gamma(\frac{t}{c})}{\Gamma(km)}\Big].$$

Finally, equation (36) reduces to

$$\begin{aligned}\mathrm{E}(X_{i:n}^t) \;=\; & t\sum_{p=n-i+1}^{n}(-1)^{p-n+i-1}\binom{p-1}{n-i}\binom{n}{p}\sum_{l=0}^{p}\sum_{j,m=0}^{\infty}\frac{(-1)^{l+m}\binom{p}{l}s^t\,c_{l,j}}{c\,m!\,(1-\mathrm{e}^{-\lambda})^1}\\ & \times\;\Big[\frac{\Gamma[(j+l)\,\alpha+1]\,\Gamma(km-\frac{t}{c})\,\Gamma(\frac{t}{c})}{\Gamma[(j+l)\,\alpha+1-m]\,\Gamma(km)}\Big].\end{aligned}$$

The L moments (Hosking, 1990) are expectations of certain linear combinations of order statistics and can be used to calculate quantities analogous to standard deviation, skewness and kurtosis, termed the L scale, L skewness and L kurtosis respectively. They are defined by

$$\lambda_{r+1} = (r+1)^{-1}\sum_{k=0}^{r}(-1)^k\binom{r}{k}E(X_{r+1-k:r+1}),\;\; r=0,1,\ldots$$

The first four L-moments are: $\lambda_1 = E(X_{1:1})$, $\lambda_2 = \frac{1}{2}E(X_{2:2}-X_{1:2})$, $\lambda_3 = \frac{1}{3}E(X_{3:3}-2X_{2:3}+X_{1:3})$ and $\lambda_4 = \frac{1}{4}E(X_{4:4}-3X_{3:4}+3X_{2:4}-X_{1:4})$. From equation (35) for the moments of the order statistics, we can obtain expansions for the L-moments of the Exp-BXIIP distribution as linear functions of the means of suitable Exp-BXII distributions.

8. Rényi and Shannon Entropy

The entropy of a random variable X with density function $f(x)$ is a measure of variation of the uncertainty. For any real parameter $\omega > 0$ and $\omega \neq 1$, the Rényi entropy of the Exp-BXIIP distribution is given by

$$\begin{aligned}I_R(\gamma) \;&=\; \frac{1}{(1-\gamma)}\log\int_0^{\infty}f^{\gamma}(x)dx\\ &=\; \frac{1}{(1-\gamma)}\log\Big\{\Big[\frac{c\,k\,s^{-c}\,\alpha\,\lambda}{1-\mathrm{e}^{-\lambda}}\Big]^{\gamma}\sum_{j,r=0}^{\infty}\frac{(-1)^{j+r}\lambda^j\,\gamma^j\binom{(j+r)\alpha-\gamma}{r}}{j!}\\ &\times\;\int_0^{\infty}x^{(c-1)\gamma}\Big[1+\Big(\frac{x}{s}\Big)^c\Big]^{-k(r+\gamma)-\gamma}dx\Big\}\\ &=\; \frac{1}{(1-\gamma)}\log\Big\{\Big[\frac{c\,k\,s^{-c}\,\alpha\,\lambda}{1-\mathrm{e}^{-\lambda}}\Big]^{\gamma}\frac{s^{(c-1)\gamma+1}}{c}\sum_{j,r=0}^{\infty}\frac{(-1)^{j+r}\lambda^j\,\gamma^j\binom{(j+r)\alpha-\gamma}{r}}{j!}\\ &\times\;\Big[\frac{\Gamma[\frac{(c-1)\gamma+1}{c}]\,\Gamma[\frac{ck(r+\gamma)+\gamma-1}{c}]}{\Gamma[k(r+\gamma)+\gamma]}\Big]\Big\}.\end{aligned}$$

The details of the proof are given in Appendix B.

For the Shannon entropy, we have

$$\begin{aligned} E\{-\log[f(X)]\} &= \log(1-\mathrm{e}^{-\lambda}) - \log(\alpha) - \log(\lambda) \\ &+ (\alpha-1)\sum_{j,r=0}^{\infty} \frac{(r+1)\,v_r}{(j+1)\,(j+r+2)} \\ &+ \lambda \sum_{j,r=0}^{\infty} \frac{(-1)^j (r+1)\,v_r\,\Gamma(\alpha+1)}{(j+r+1)\,j!\,\Gamma(\alpha-j)} - E\{\log[g(X;s,k,c)]\}, \end{aligned}$$

where v_r is defined in Theorem 1 and $E\{\log[g(X;s,k,c)]\}$ can be computed from (2) at least numerically. The details of the proof are given in Appendix C.

9. Estimation

Let X_i be a random variable following (9) with the vector $\boldsymbol{\theta} = (s,k,c,\alpha,\lambda)^T$ of parameters. The data encountered in survival analysis and reliability studies are often censored. The censored log-likelihood $l(\boldsymbol{\theta})$ for the model parameters is

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log(c) + \log(k) - c\log(s) + \log(\alpha) + \log(\lambda) - \log(1-\mathrm{e}^{-\lambda}) \\ &+ \frac{(c-1)}{n}\sum_{i=1}^{n}\log(x_i) + \frac{k}{n}\sum_{i=1}^{n}\log(q_{i,-1}) + \frac{(\alpha-1)}{n}\sum_{i=1}^{n}\log(v_{i,1}) \\ &- \frac{\lambda}{n}\sum_{i=1}^{n} v_{i,\alpha}. \end{aligned}$$

The score functions for the parameters s, k, c α and λ are given by

$$U_s(\boldsymbol{\theta}) = -\frac{c}{s} - \frac{kc}{ns}\sum_{i=1}^{n} u_i\,q_{i,1} - \frac{ck(\alpha-1)}{ns}\sum_{i=1}^{n}\frac{u_i\,q_{i,k+1}}{v_{i,1}},$$

$$U_k(\boldsymbol{\theta}) = \frac{1}{k} + \frac{1}{n}\sum_{i=1}^{n}\log(q_{i,-1}) + \frac{(\alpha-1)}{n}\sum_{i=1}^{n}\frac{q_{i,k}\,\log(q_{i,-1})}{v_{i,1}},$$

$$U_c(\boldsymbol{\theta}) = \frac{1}{c} - \log(s) + \frac{1}{n}\sum_{i=1}^{n}\log(x_i) + \frac{k}{n}\sum_{i=1}^{n}\frac{u_i\,\log(u_i^{1/c})}{q_{i,-1}},$$

$$U_\alpha(\boldsymbol{\theta}) = \frac{1}{\alpha} + \frac{1}{n}\sum_{i=1}^{n}\log(v_{i,1}) - \frac{\lambda}{n}\sum_{i=1}^{n} v_{i,\alpha}\,\log(v_{i,1}),$$

and

$$U_\lambda(\boldsymbol{\theta}) = \frac{1}{\lambda} - \frac{\mathrm{e}^{-\lambda}}{1-\mathrm{e}^{-\lambda}} - \sum_{i=1}^{n} v_{i,\alpha},$$

where $u_i = \left(\frac{x_i}{s}\right)^c$, $q_{i,k} = \left[1+\left(\frac{x_i}{s}\right)^c\right]^{-k}$ and $v_{i,\alpha} = \left\{1-\left[1+\left(\frac{x_i}{s}\right)^c\right]^{-k}\right\}^\alpha$.

The maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by solving the nonlinear likelihood equations $U_s(\boldsymbol{\theta}) = 0$, $U_k(\boldsymbol{\theta}) = 0$, $U_c(\boldsymbol{\theta}) = 0$, $U_\alpha(\boldsymbol{\theta}) = 0$ and $U_\lambda(\boldsymbol{\theta}) = 0$. These equations cannot be solved analytically and statistical software can be used to solve them numerically. We can use iterative techniques such as a Newton-Raphson type algorithm to obtain $\widehat{\boldsymbol{\theta}}$. The computations are performed using the software R version 3.0.0 (package bbmle).

For interval estimation of (s,k,c,α,λ) and hypothesis tests on these parameters, we obtain the observed information matrix since its expectation requires numerical integration. The 5×5 observed information matrix $J(\boldsymbol{\theta})$ is

$$J(\boldsymbol{\theta}) = -\begin{pmatrix} \mathbf{U}_{ss} & \mathbf{U}_{sk} & \mathbf{U}_{sc} & \mathbf{U}_{s\alpha} & \mathbf{U}_{s\lambda} \\ . & \mathbf{U}_{kk} & \mathbf{U}_{kc} & \mathbf{U}_{k\alpha} & \mathbf{U}_{k\lambda} \\ . & . & \mathbf{U}_{cc} & \mathbf{U}_{c\alpha} & \mathbf{U}_{c\lambda} \\ . & . & . & \mathbf{U}_{\alpha\alpha} & \mathbf{U}_{\alpha\lambda} \\ . & . & . & & \mathbf{U}_{\lambda\lambda} \end{pmatrix},$$

whose elements are given in Appendix D. The matrix $J(\boldsymbol{\theta})$ is useful to obtain approximate confidence intervals for the parameters.

10. Application

In this section, we illustrate the usefulness of the Exp-BXIIP distribution applied to a real data set. These data on failure times are reported in the book "Weibull Models by Murthy" *et al.* (2004, page 297). We also fit the density functions of the *Exponentiated Burr XII Poisson* (Exp-BXIIP), *Beta Burr XII* (BBXII), *Kumaraswamy Burr XII* (KwBXII) and *McDonald Burr XII* (McBXII) distributions given by

$$\begin{aligned} f_{\text{Exp-BXIIP}}(x;s,k,c,\alpha,\lambda) &= \frac{c\,k\,s^{-c}\,\alpha\,\lambda}{1-\mathrm{e}^{-\lambda}}\,x^{c-1}\Big[1+\Big(\frac{x}{s}\Big)^{c}\Big]^{-k-1}\Big\{1-\Big[1+\Big(\frac{x}{s}\Big)^{c}\Big]^{-k}\Big\}^{\alpha-1} \\ &\times \exp\{-\lambda\Big[1-(1+(\frac{x}{s})^{c}\Big)^{-k}\Big]^{\alpha}\}, \end{aligned}$$

$$f_{\text{BBXII}}(x;s,k,c,a,b) = \frac{c\,k\,s^{-c}}{B(a,b)}\,x^{c-1}\Big[1+\Big(\frac{x}{s}\Big)^{c}\Big]^{-k\,b-1}\Big\{1-\Big[1+\Big(\frac{x}{s}\Big)^{c}\Big]^{-k}\Big\}^{a-1},$$

$$\begin{aligned} f_{\text{KwBXII}}(x;s,k,c,a,b) &= a\,b\,c\,k\,s^{-c}\,x^{c-1}\Big[1+\Big(\frac{x}{s}\Big)^{c}\Big]^{-k-1}\Big\{1-\Big[1+\Big(\frac{x}{s}\Big)^{c}\Big]^{-k}\Big\}^{\alpha-1} \\ &\times \left\{1-\Big\{1-\Big[1+\Big(\frac{x}{s}\Big)^{c}\Big]^{-k}\Big\}^{a}\right\}^{b-1}, \end{aligned}$$

$$\begin{aligned} f_{\text{McBXIIP}}(x;s,k,c,a,b,\alpha) &= \frac{c\,k\,s^{-c}\,\alpha}{B(a,b)}\,x^{c-1}\Big[1+\Big(\frac{x}{s}\Big)^{c}\Big]^{-k-1}\Big\{1-\Big[1+\Big(\frac{x}{s}\Big)^{c}\Big]^{-k}\Big\}^{a\,\alpha-1} \\ &\times \left\{1-\Big\{1-\Big[1+\Big(\frac{x}{s}\Big)^{c}\Big]^{-k}\Big\}^{\alpha}\right\}^{b-1}, \end{aligned}$$

respectively, where all parameters are positive.

Further, we apply the Cramér-von Mises (W^*) and Anderson-Darling (A^*) statistics described in details in Chen and Balakrishnan (1995) to verify which distribution fits better to these data. In general, the smaller the values of the statistics W^* and A^*, the better the fit to the data. Let $H(x;\boldsymbol{\theta})$ be the cdf, where the form of H is known but $\boldsymbol{\theta}$ (a k-dimensional parameter vector, say) is unknown. To obtain the statistics W^* and A^*, one can proceed as follows: (i) Compute $v_i = H(x_i;\widehat{\boldsymbol{\theta}})$, where the x_i's are in ascending order; (ii) Compute $y_i = \Phi^{-1}(v_i)$, where $\Phi(\cdot)$ is the standard normal cdf and $\Phi^{-1}(\cdot)$ its inverse; (iii) Compute $u_i = \Phi\{(y_i-\bar{y})/s_y\}$, where $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$ and $s_y^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$; (iv) Calculate $W^2 = \sum_{i=1}^{n}\{u_i-(2i-1)/(2n)\}^2 + 1/(12n)$ and $A^2 = -n-(1/n)\sum_{i=1}^{n}\{(2i-1)\log(u_i)+(2n+1-2i)\log(1-u_i)\}$; (v) Modify W^2 into $W^* = W^2(1+0.5/n)$ and A^2 into $A^* = A^2(1+0.75/n+2.25/n^2)$. Table 1 and 2, respectively, lists the MLEs, their standard errors in parentheses and the statistics W^* and A^* and p-values for the failure times data. They indicate that the Exp-BXIIP and McBXII distributions are the best models to these data. Morever, the standard errors are much smaller compared with their estimates for the Exp-BXIIP distribution.

Table 1. MLEs

Distribution	Estimatives					
Exp-BXIIP	$\widehat{s}$	$\widehat{k}$	$\widehat{c}$	$\widehat{\alpha}$	$\widehat{\lambda}$	
	14.8518	5.9646	5.3264	0.4471	22.7252	
	(0.0726)	(0.2732)	(0.0734)	(0.0427)	(5.9532)	
KwBXII	$\widehat{s}$	$\widehat{k}$	$\widehat{c}$	$\widehat{a}$	$\widehat{b}$	
	5.7916	6.3749	7.0604	0.2510	1.4676	
	(1.3907)	(5.8618)	(0.0207)	(0.0584)	(0.9002)	
McBXII	$\widehat{s}$	$\widehat{k}$	$\widehat{c}$	$\widehat{a}$	$\widehat{b}$	$\widehat{\alpha}$
	6.5058	6.7187	6.6588	0.5303	2.3546	0.5081
	(0.0103)	(6.6788)	(0.0103)	(0.6782)	(1.8471)	(0.6112)
BBXII	$\widehat{s}$	$\widehat{k}$	$\widehat{c}$	$\widehat{a}$	$\widehat{b}$	
	7.5361	6.5139	6.3234	0.2584	6.4360	
	(0.0553)	(9.2920)	(0.0599)	(0.0357)	(9.1064)	

Table 2. Measures W^* and A^*

Distribution	W^*	p-value	A^*	p-value
Exp-BXIIP	0.05853	0.395	0.58938	0.124
KwBXII	0.09915	0.115	0.68478	0.074
McBXII	0.10423	0.098	0.95368	0.016
BBXII	0.13694	0.035	1.19636	0.004

More information is provided by a visual comparison of the fitted densities to the histogram of the data. The plots of the fitted Exp-BXIIP, BBXII, KwBXII and McBXII density functions are displayed in Figure 4. These plots indicate that the new distribution provides a good fit to these data and that it is also a very competitive model to other lifetime distributions.

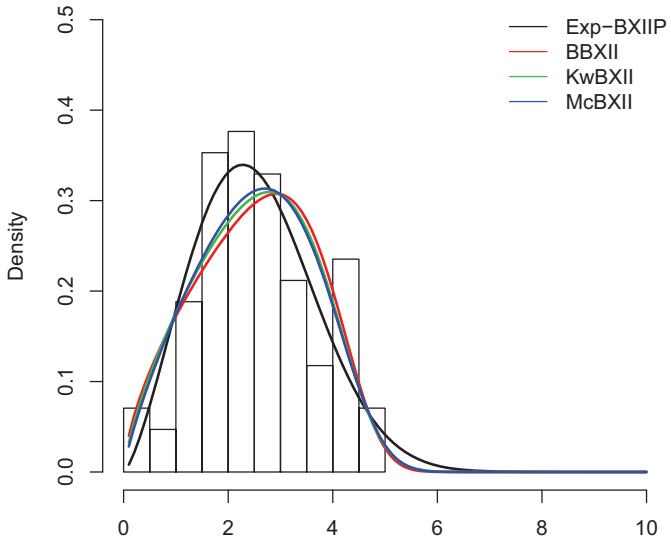


Figure 4. Fitted densities to the histogram of the current data.

11. Conclusions

We define and study a new five-parameter lifetime model called the *exponentiated Burr XII Poisson* distribution, which extends some well-known lifetime distributions. Due to its flexibility in accommodating different forms of the hazard rate function, it is an important model for modeling lifetime data. We provide a mathematical treatment of the proposed distribution including a useful expansion for its density function. We derive explicit expressions for the moments, generating and quantile functions, mean deviations, reliability and entropies, which hold in generality for any parameter values. The model parameters are estimated by maximum likelihood. Additionally, the observed information matrix is determined. In one application to a real data set, we illustrate the potentiality of the new model.

Appendix A - Generating function

We have the following result which holds for m and k positive integers, $\mu > -1$ and $p > 0$ (Prudnikov et al., 1992,

page 21)

$$
\begin{aligned}
I\left(p,\mu,\frac{m}{k},\nu\right) &= \int_0^\infty \exp(-px)\, x^{\mu}\,(1+x^{\frac{m}{k}})^{\nu} dx \\
&= \frac{k^{-\nu} m^{\mu+\frac{1}{2}}}{(2\pi)^{\frac{(m-1)}{2}}\,\Gamma(-\nu)\, p^{\mu+1}} \times \\
&\quad G^{k,k+m}_{k+m,k}\left(\frac{m^m}{p^m}\,\middle|\,\begin{matrix}\Delta(m,-\mu),\Delta(k,\nu+1)\\ \Delta(k,0)\end{matrix}\right),
\end{aligned} \tag{38}
$$

where $\Delta(k,a) = \frac{a}{k}, \frac{a+1}{k}, \cdots, \frac{a+k}{k}$.

Appendix B - Rényi entropy

The entropy of a random variable X with density function $f(x)$ is a measure of variation of the uncertainty. For any real parameter $\omega > 0$ and $\omega \neq 1$, the Rényi entropy is given by

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

where

$$
\begin{aligned}
f(x)^{\gamma} &= \left[\frac{c\,k\,s^{-c}\,\alpha\,\lambda}{1-\mathrm{e}^{-\lambda}}\right]^{\gamma} x^{(c-1)\gamma}\left[1+\left(\frac{x}{s}\right)^{c}\right]^{-(k+1)\gamma}\left\{1-\left[1+\left(\frac{x}{s}\right)^{c}\right]^{-k}\right\}^{(\alpha-1)\gamma} \\
&\times \exp\left\{-\lambda\,\gamma\left\{1-\left[1+\left(\frac{x}{s}\right)^{c}\right]^{-k}\right\}^{\alpha}\right\} \\
&= \left[\frac{c\,k\,s^{-c}\,\alpha\,\lambda}{1-\mathrm{e}^{-\lambda}}\right]^{\gamma} x^{(c-1)\gamma}\left[1+\left(\frac{x}{s}\right)^{c}\right]^{-(k+1)\gamma} \sum_{j=0}^{\infty} \frac{(-1)^j\,\lambda^j\,\gamma^j}{j!} \\
&\times \left\{1-\left[1+\left(\frac{x}{s}\right)^{c}\right]^{-k}\right\}^{(j+\gamma)\,\alpha-\gamma} \\
&= \left[\frac{c\,k\,s^{-c}\,\alpha\,\lambda}{1-\mathrm{e}^{-\lambda}}\right]^{\gamma} x^{(c-1)\gamma} \sum_{j,r=0}^{\infty} \frac{(-1)^{j+r}\,\lambda^j\,\gamma^j\binom{(j+\gamma)\,\alpha-\gamma)}{r}}{j!}\left[1+\left(\frac{x}{s}\right)^{c}\right]^{-k(r+\gamma)-\gamma}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
I_R(\gamma) &= \frac{1}{(1-\gamma)}\log\left\{\left[\frac{c\,k\,s^{-c}\,\alpha\,\lambda}{1-\mathrm{e}^{-\lambda}}\right]^{\gamma} \sum_{j,r=0}^{\infty} \frac{(-1)^{j+r}\lambda^j\,\gamma^j\binom{(j+r)\alpha-\gamma}{r}}{j!}\right. \\
&\quad \left.\times \int_0^\infty x^{(c-1)\gamma}\left[1+\left(\frac{x}{s}\right)^{c}\right]^{-k(r+\gamma)-\gamma} dx\right\} \\
&= \frac{1}{(1-\gamma)}\log\left\{\left[\frac{c\,k\,s^{-c}\,\alpha\,\lambda}{1-\mathrm{e}^{-\lambda}}\right]^{\gamma} \frac{s^{(c-1)\gamma+1}}{c} \sum_{j,r=0}^{\infty} \frac{(-1)^{j+r}\lambda^j\,\gamma^j\binom{(j+r)\alpha-\gamma}{r}}{j!}\right. \\
&\quad \left.\times\left[\frac{\Gamma[\frac{(c-1)\gamma+1}{c}]\,\Gamma[\frac{ck(r+\gamma)+\gamma-1}{c}]}{\Gamma[k(r+\gamma)+\gamma]}\right]\right\}.
\end{aligned}
$$

Appendix C - Shannon entropy

The Shannon entropy of a random variable X with density function $f(x)$ is a measure of variation of the uncertainty. The Shannon entropy is given by

$$
\begin{aligned}
E\{-\log[f(X)]\} &= \log(1-\mathrm{e}^{-\lambda}) - \log(\alpha) - \log(\lambda) - (\alpha-1)\,\mathrm{E}\{\log[\mathrm{G}(\mathrm{X})]\} \\
&+ \lambda\, E[G^{\alpha}(X)] - E\{\log[g(X)]\} \\
&= \log(1-\mathrm{e}^{-\lambda}) - \log(\alpha) - \log(\lambda) + (\alpha-1)\sum_{j,r=0}^{\infty} \frac{(r+1)\,v_r}{(j+1)\,(j+r+2)} \\
&+ \lambda \sum_{j,r=0}^{\infty} \frac{(-1)^j (r+1)\,v_r\,\Gamma(\alpha+1)}{(j+r+1)\,j!\,\Gamma(\alpha-j)} - E\{\log[g(X)]\}.
\end{aligned}
$$

We use the power series

$$G^{\alpha}(X) = \left\{1 - \left[1 + \left(\frac{x}{s}\right)^{c}\right]^{-k}\right\}^{\alpha} = \sum_{j=0}^{\infty} \frac{(-1)^{j}\,\Gamma(\alpha+1)}{j!\,\Gamma(\alpha-j)} \left[1 + \left(\frac{x}{s}\right)^{c}\right]^{-jk}$$

and

$$\begin{aligned} \log[G(X)] &= \sum_{j=0}^{\infty} \frac{(-1)^{j}}{j+1} \left[G(x) - 1\right]^{j+1} = \sum_{j=0}^{\infty} \frac{(-1)^{j}}{j+1} \left\{ - \left[1 + \left(\frac{x}{s}\right)^{c}\right]^{-k}\right\}^{j+1} \\ &= \sum_{j=0}^{\infty} \frac{(-1)^{2j+1}}{j+1} \left[1 + \left(\frac{x}{s}\right)^{c}\right]^{-k(j+1)} = - \sum_{j=0}^{\infty} \frac{1}{j+1} \left[1 + \left(\frac{x}{s}\right)^{c}\right]^{-k(j+1)}. \end{aligned}$$

Next, we have

$$\begin{aligned} E[G^{\alpha}(X)] &= \sum_{r=0}^{\infty} v_r \int_0^{\infty} G^{\alpha}(x)\, g(x; s, k(r+1), c) dx \\ &= c\,k\,s^{-c}\,\Gamma(\alpha+1) \sum_{j,r=0}^{\infty} \frac{(-1)^{j}\,(r+1)\,v_r}{j!\,\Gamma(\alpha-j)} \int_0^{\infty} x^{c-1} \left[1 + \left(\frac{x}{s}\right)^{c}\right]^{-k(j+r+1)-1} dx \\ &= c\,k\,s^{-c}\,\Gamma(\alpha+1) \sum_{j,r=0}^{\infty} \frac{(-1)^{j}\,(r+1)\,v_r}{j!\,\Gamma(\alpha-j)} \frac{s^{c}}{c} \int_0^{\infty} (1+u)^{-k(j+r+1)-1} dx \\ &= \sum_{j,r=0}^{\infty} \frac{(-1)^{j}\,\Gamma(\alpha+1)\,(r+1)\,v_r \Gamma(\alpha+1)}{(j+r+1)\,j!\,\Gamma(\alpha-j)} \end{aligned}$$

and

$$\begin{aligned} E\{\log[G(X)]\} &= \sum_{r=0}^{\infty} v_r \int_0^{\infty} \log G(x)\, g(x; s, k(r+1), c) dx \\ &= -c\,k\,s^{-c} \sum_{j,r=0}^{\infty} \frac{(r+1)\,v_r}{j+1} \int_0^{\infty} x^{c-1} \left[1 + \left(\frac{x}{s}\right)^{c}\right]^{-k(j+r+2)-1} dx \\ &= -c\,k\,s^{-c} \sum_{j,r=0}^{\infty} \frac{(r+1)\,v_r}{j+1} \frac{s^{c}}{c} \int_0^{\infty} (1+u)^{-k(j+r+2)-1} dx \\ &= - \sum_{j,r=0}^{\infty} \frac{(r+1)\,v_r}{(j+1)(j+r+2)}. \end{aligned}$$

Thus,

$$\begin{aligned} E\{-\log[f(X)]\} &= \log(1 - \mathrm{e}^{-\lambda}) - \log(\alpha) - \log(\lambda) + (\alpha-1) \sum_{j,r=0}^{\infty} \frac{(r+1)\,v_r}{(j+1)\,(j+r+2)} \\ &+ \lambda \sum_{j,r=0}^{\infty} \frac{(-1)^{j}(r+1)\,v_r\,\Gamma(\alpha+1)}{(j+r+1)\,j!\,\Gamma(\alpha-j)} - E[\log g(X)]. \end{aligned}$$

Appendix D - Information matrix

The elements of the observed information matrix $J(\boldsymbol{\theta})$ for the parameters $(s, k, c, \alpha, \lambda)$ are

$$
\begin{aligned}
U_{ss}(\boldsymbol{\theta}) &= \frac{c}{s^2} + \frac{kc}{ns^2}\sum_{i=1}^{n}\frac{u_i^2\,(c+q_{i,-1})}{q_{i,-2}} \\
&+ \frac{kc(\alpha-1)}{ns^2}\sum_{i=1}^{n}\frac{u_i\Big[c\Big(1+k\,u_i\,q_{i,-k}-q_{i,-k}\Big)-q_{i,-1}(q_{i,-k}-1)\Big]}{q_{i,-2}\,(q_{i,-k}-1)^2} \\
&- \frac{kc\alpha\lambda}{ns^2}\sum_{i=1}^{n}\frac{u_i\,v_{i,\alpha}\Big\{c\Big[-k\,u_i(q_{i,-k}-\alpha)+(q_{i-k}-1)\Big]+q_{i,-1}(q_{i,-k}-1)\Big\}}{q_{i,-2}\,(q_{i,-k}-1)^2},
\end{aligned}
$$

$$
\begin{aligned}
U_{sk}(\boldsymbol{\theta}) &= -\frac{c}{ns^2}\sum_{i=1}^{n}\frac{x_i\,u_i^{(c-1)/c}}{q_{i,-1}} + \frac{c(\alpha-1)}{ns}\sum_{i=1}^{n}\frac{u_i\Big[1-q_{i,-k}+k\,q_{i,-k}\log(q_{i,-1})\Big]}{q_{i,-1}\,(q_{i,-k}-1)^2} \\
&- \frac{c\alpha\lambda}{ns}\sum_{i=1}^{n}\frac{u_i\,v_{i,\alpha}\Big[-k\log(q_{i,-1})u_i(q_{i,-k}-\alpha)+q_{i-k}-1\Big]}{q_{i,-1}\,(q_{i,-k}-1)^2},
\end{aligned}
$$

$$
\begin{aligned}
U_{sc}(\boldsymbol{\theta}) &= -\frac{1}{s} - \frac{k}{ns}\sum_{i=1}^{n}\frac{u_i\Big[q_{i,-1}+c\log(u_i^{1/c})\Big]}{q_{i,-2}} \\
&+ \frac{k(\alpha-1)}{ns}\sum_{i=1}^{n}\frac{u_i\Big\{c\log(u_i^{1/c})\Big[1+k\,u_i\,q_{i,-k}-q_{i,-k}\Big]-q_{i,-1}(q_{i,-k}-1)\Big\}}{q_{i,-2}\,(q_{i,-k}-1)^2} \\
&- \frac{kc\alpha\lambda}{ns}\sum_{i=1}^{n}\frac{u_i\,v_{i,\alpha}\Big\{\log(u_i^{1/c})\Big[k\,u_i(q_{i,-k}-\alpha)-v_{i,1}-q_{i,-1}(q_{i,-k}-1)\Big]\Big\}}{q_{i,-2}\,(q_{i,-k}-1)^2},
\end{aligned}
$$

$$
U_{s\alpha}(\boldsymbol{\theta}) = -\frac{kc}{ns^2}\sum_{i=1}^{n}\frac{x_i\,u_i^{(c-1)/c}\,q_{i,k+1}}{v_{i,1}} + \frac{kc\lambda}{ns}\sum_{i=1}^{n}\frac{u_i\,v_{i,\alpha}\Big[1+\alpha\,\log(v_{i,1})\Big]}{q_{i,-1}\,(q_{i,-k}-1)},
$$

$$
U_{s\lambda}(\boldsymbol{\theta}) = \frac{kc\alpha}{ns}\sum_{i=1}^{n}\frac{u_i\,q_{i,k+1}}{v_{i,\alpha-1}},
$$

$$
U_{kk}(\boldsymbol{\theta}) = -\frac{1}{k^2} - \frac{(\alpha-1)}{n}\sum_{i=1}^{n}\frac{q_{i,-k}\,\log^2(q_{i,-1})}{(q_{i,-k}-1)^2} + \frac{\alpha\lambda}{n}\sum_{i=1}^{n}\frac{v_{i,\alpha}\,(q_{i,-k}-\alpha)\,\log^2(q_{i,-1})}{(q_{i,-k}-1)^2},
$$

$$
\begin{aligned}
U_{kc}(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\frac{u_i\,\log(u_i^{1/c})}{q_{i,-1}} - \frac{(\alpha-1)}{n}\sum_{i=1}^{n}\frac{u_i\,\log(u_i^{1/c})\Big[1-q_{i,-k}+k\,q_{i,-k}\log(q_{i,-1})\Big]}{q_{i,-1}\,(q_{i,-k}-1)^2} \\
&+ \frac{k\alpha\lambda}{n}\sum_{i=1}^{n}\frac{u_i\,\log(u_i^{1/c})\,v_{i,\alpha}\,\log(q_{i,-1})}{q_{i,-1}\,(q_{i,-k}-1)^2},
\end{aligned}
$$

$$
U_{k\alpha}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{q_{i,k}\,\log(q_{i,-1})}{v_{i,1}} - \frac{\lambda}{n}\sum_{i=1}^{n}\frac{\log(q_{i,-1})\,v_{i,\alpha}\Big[1+\alpha\,\log(v_{i,1})\Big]}{q_{i,-k}-1},
$$

$$U_{k\lambda}(\boldsymbol{\theta}) = -\frac{\alpha}{n}\sum_{i=1}^{n} q_{i,k}\, v_{i,\alpha}\, \log(q_{i,-1}),$$

$$\begin{aligned} U_{cc}(\boldsymbol{\theta}) &= -\frac{1}{c^2} + \frac{k}{n}\sum_{i=1}^{n}\frac{u_i\, \log^2(u_i^{1/c})}{q_{i,-2}} \\ &- \frac{k(\alpha-1)}{n}\sum_{i=1}^{n}\frac{u_i\, \log^2(u_i^{1/c})\Big[1 + k\, u_i\, q_{i,-k} - q_{i,-k}\Big]}{q_{i,-2}\,(q_{i,-k}-1)^2} \\ &- \frac{k\alpha\lambda}{n}\sum_{i=1}^{n}\frac{u_i\, \log^2(u_i^{1/c})\, v_{i,\alpha}\Big[-k\, u_i(q_{i,-k}-\alpha) + q_{i-k} - 1\Big]}{q_{i,-2}\,(q_{i,-k}-1)^2}, \end{aligned}$$

$$U_{c\alpha}(\boldsymbol{\theta}) = \frac{k}{n}\sum_{i=1}^{n}\frac{u_i\, q_{i.k+1}\, \log(u_i^{1/c})}{v_{i,1}} - \frac{k\lambda}{n}\sum_{i=1}^{n}\frac{u_i\, \log(u_i^{1/c})\, v_{i,\alpha}\Big[1 + \alpha\, \log(v_{i,1})\Big]}{q_{i,-1}\,(q_{i,-k}-1)},$$

$$U_{c\lambda}(\boldsymbol{\theta}) = -\frac{k\alpha}{n}\sum_{i=1}^{n} u_i\, q_{i,k+1}\, v_{i,\alpha-1}\, \log(u_i^{1/c}),$$

$$U_{\alpha\alpha}(\boldsymbol{\theta}) = -\frac{1}{\alpha^2} - \frac{\lambda}{n}\sum_{i=1}^{n} v_{i,\alpha}\, \log^2(v_{i,1}),$$

$$U_{\alpha\lambda}(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n} v_{i,\alpha}\, \log(v_{i,1}),$$

$$U_{\lambda\lambda}(\boldsymbol{\theta}) = -\frac{1}{\lambda^2} + \frac{\mathrm{e}^{\lambda}}{(\mathrm{e}^{\lambda}-1)^2},$$

where u_i, $q_{i,k}$, $v_{i,\alpha}$ are given in Section 9.

References

Brito, E., Silva, G. O., Cordeiro, G. M., & Demétrio, C. G. B. (2014). The gamma Burr XII distribution: Theory and Practice. *Journal of Data Science, In Press*.

Chen, G., & Balakrishnan, N. (1995). A general purpose approximate goodness-of-fit test. *Journal of Quality Technology, 27*, 154-161.

Cordeiro, G. M., Ortega, E. M. M., Hamedani, G. G., & Garcia, D. A. (2012). The McBurr XII Model. *Accepted*.

El-Bassiouny, A. H., & Abdo, N. F. (2010). Reliability properties of seven parameters Burr XII distribution. *Comput Meth Sci Tech, 16*, 127-133. http://dx.doi.org/10.12921/cmst.2010.16.02.127-133

Gradshteyn, I. S., & Ryzhik, I. M. (2000). *Table of Integrals, Series and Products*. Academic Press, San Diego.

Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society B, 52*, 105-124.

Jayakumar, K., & Mathew, T. (2008). On a generalization to Marshall-Olkin scheme and its application to Burr type XII distribution. *Stat Pap, 49*, 421-439. http://dx.doi.org/10.1007/s00362-006-0024-5

Kenney, J. F., & Keeping, E. S. (1962). *Mathematics of Statistics*, 3rd ed., Part 1. New Jersey.

Murthy, D. N. P., Xie. M., & Jiang, R. (2004). *Weibull Models*, John Wiley and Sons, New Jersey.

Parnaiba, P. F. P., Ortega, E. M. M., Cordeiro, G. M., & Pescim, R. R. (2011). The beta Burr XII distribution with application to lifetime data. *Computational Statistics and Data Analysis, 55*, 1118-1136. http://dx.doi.org/10.1016/j.csda.2010.09.009

Parnaiba, P. F. P., Ortega, E. M. M., Cordeiro, G. M., & Pascoa, M. A. R. (2012). The Kumaraswamy Burr XII distribution: Theory and Practice. *Journal of Statistical Computation and Simulation, 82*, 1-27.

Prudnikov, A. P., Brychkov, Y. A., & Marichev, O.I. (1986). *Integrals and Series*, volume 1. Gordon and Breach Science Publishers, Amsterdam.

Prudnikov, A. P., Brychkov, Y. A., & Marichev, O. I. (1992). *Integrals and Series*, volume 4. Gordon and Breach Science Publishers, Amsterdam.

Ramos, M. W. A., Percontini, A., Cordeiro, G. M., & Silva, R. V. (2015). The Burr XII Negative Binomial Distribution with Applications to Lifetime Data. *International Journal of Statistics and Probability, 4*, 109-125. http://dx.doi.org/10.5539/ijsp.v4n1p109

R Development Core Team. (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna.

Rényi, A. (1961). On measures of entropy and information. *In: Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Volume I, pp. 547–561*. University of California Press: Berkeley.

Ristic, M. M., & Nadarajah, S. (2012). A new lifetime distribution. *Journal of Statistical Computation and Simulation, 81*, 1-16.

Shao, Q. (2004a). Notes on maximum likelihood estimation for the three-parameter Burr XII distribution . *Computational Statistics and Data Analysis, 45*, 675-687. http://dx.doi.org/10.1016/S0167-9473(02)00367-5

Shao, Q., Wong, H., & Xia, J. (2004b). Models for extremes using the extended three parameter Burr XII system with application to flood frequency analysis. *Hydrological Sciences Journal des Sciences Hydrologiques, 49*, 685-702. http://dx.doi.org/10.1623/hysj.49.4.685.54425

Soliman, A. A. (2005). Estimation of Parameters of Life From Progressively Censored Data Using Burr-XII Model. *IEEE Transactions on Reliability, 54*, 34-42. http://dx.doi.org/10.1109/TR.2004.842528

Zimmer, W. J., Keats, J. B., & Wang, F. K. (1998). The Burr XII distribution in reliability analysis. *Journal of Quality Technology, 30*, 386-94.

International Journal of Statistics and Probability; Vol. 4, No. 4; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

The Transmuted Marshall-Olkin Fréchet Distribution: Properties and Applications

Ahmed Z. Afify[1], G. G. Hamedani[2], Indranil Ghosh[3], & M. E. Mead[4]

[1] Department of Statistics, Mathematics and Insurance, Benha University, Egypt

[2] Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, USA

[3] Department of Mathematics and Statistics, University of North Carolina Wilmington, USA

[4] Department of Statistics and Insurance, Faculty of Commerce, Zagazig University, Egypt

Correspondence: Indranil Ghosh, Department of Mathematics & Statistics, University of North Carolina Wilmington, Wilmington, USA. Tel: 1-910-962-7644. E-mail: ghoshi@uncw.edu

Received: September 12, 2015 Accepted: September 28, 2015 Online Published: October 26, 2015

doi:10.5539/ijsp.v4n4p132 URL: http://dx.doi.org/10.5539/ijsp.v4n4p132

Abstract

This paper introduces a new four-parameter lifetime model, which extends the Marshall-Olkin Fréchet distribution introduced by Krishna et al. (2013), called the transmuted Marshall-Olkin Fréchet distribution. Various structural properties including ordinary and incomplete moments, quantile and generating function, Rényi and q-entropies and order statistics are derived. The maximum likelihood method is used to estimate the model parameters. We illustrate the superiority of the proposed distribution over other existing distributions in the literature in modeling two real life data sets.

Keywords: Transmuted family, generating Function, Rényi Entropy, order Statistics, maximum Likelihood estimation, Marshall-Olkin Fréchet distribution.

1. Introduction

Recently, there has been an increased interest in developing generalized continuous univariate distributions which have been extensively used for analyzing and modeling data in many applied areas such as lifetime analysis, engineering, economics, insurance and environmental sciences. However, these applied areas clearly require extended forms of these probability distributions when the parent models do not provide adequate fits. So, several families of distributions have been proposed by extending common families of continuous distributions. These generalized distributions provide more flexibility by adding one or more parameters to the baseline model. One example is the Marshal-Olkin-G (MO-G in short) family proposed by Marshal and Olkin (1997) by adding one parameter to the reliability function (rf) $\overline{G}(x) = 1 - G(x)$, where $G(x)$ is the baseline cumulative distribution function (cdf). Using the MO-G family, Krishna et al. (2013) defined and studied the Marshall-Olkin Fréchet (MOFr) distribution extending the Fréchet distribution.

The cdf of the MOFr is given (for $x > 0$) by

$$G(x, \alpha, \beta, \sigma) = \frac{e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}, \tag{1}$$

where $\sigma > 0$ is a scale parameter and α and β are positive shape parameters.

The corresponding probability density function (pdf) is given by

$$g(x, \alpha, \beta, \sigma) = \frac{\alpha\beta\sigma^{\beta} x^{-(\beta+1)} e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{\left(\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right)^2}. \tag{2}$$

The Fréchet distribution is one of the important distributions in extreme value theory and has applications in life testing, floods, rainfall, wind speeds, sea waves and track race records. Further details were explored by Kotz and Nadarajah (2000). Many authors constructed generalizations of the Fréchet distribution. For example, Nadarajah and Kotz (2003) studied the exponentiated Fréchet (EFr), Nadarajah and Gupta (2004) and Barreto-Souza et al.

(2011) independently introduced the beta Fréchet (BFr), Mahmoud and Mandouh (2013) proposed the transmuted Fréchet (TFr), Silva et al. (2013) proposed the gamma extended Fréchet (GEFr), Elbatal et al. (2014) studied the transmuted exponentiated Fréchet (TEFr), Mead and Abd-Eltawab (2014) introduced the Kumaraswamy Fréchet (Kw-Fr) and Afify et al. (2015) proposed the Weibull Fréchet (WFr) distributions.

In this paper, we define and study a new model by adding one parameter in equation (1) to provide more flexibility to the generated model. In fact, based on the transmuted-*G* (T-G) family pioneered by Shaw and Buckley (2007), we construct a new distribution called the *transmuted Marshall-Olkin Fréchet* (henceforth in short TMOFr) distribution and provide a comprehensive description of some of its mathematical properties. We hope that the new model will attract wider applications in reliability, engineering and other areas of research.

Recently, many authors used the T-G family to propose new generalizations of some well- known distributions. For example, Aryal and Tsokos (2009) defined the transmuted generalized extreme value, Aryal and Tsokos (2011) proposed the transmuted Weibull, Khan and King (2013) introduced the transmuted modified Weibull, Afify et al. (2014) defined the transmuted complementary Weibull geometric and Afify et al. (2015) proposed the transmuted Weibull Lomax distributions. For a detailed study on the general properties of the transmuted family of distributions, the interested reader is referred to Bourguignon, Ghosh and Cordeiro (2015).

Consider a baseline cdf $G(x)$ and pdf $g(x)$. Then, the cdf and pdf of the T-G family of distributions are ,respectively, defined by

$$F(x;\lambda) = G(x)\left[1+\lambda-\lambda G(x)\right] \tag{3}$$

and

$$f(x;\lambda) = g(x)\left[1+\lambda-2\lambda G(x)\right], \tag{4}$$

where $|\lambda| \leq 1$.

Note that if $\lambda = 0$, equation (4) gives the baseline distribution. Further details can be found in Shaw and Buckley (2007).

The rest of the paper is outlined as follows. In Section 2, we define the TMOFr distribution and give some plots for its pdf and hazard rate function (hrf). We derive useful mixture representations for the pdf and cdf in Section 3. We provide in Section 4 some mathematical properties of the TMOFr distribution including, ordinary and incomplete moments, moments of the residual life, reversed residual life, quantile and generating functions and Rényi and q-entropies. In Section 5, the order statistics and their moments are determined. Certain characterizations are presented in Section 6. The maximum likelihood estimates (MLEs) of the model parameters are obtained in Section 7. In Section 8, the TMOFr distribution is applied to two real data sets to illustrate its potentiality. Finally, in Section 9, we provide some concluding remarks.

2. The TMOFr Model

By inserting (1) into (3), we obtain the cdf of TMOFr (for $x > 0$)

$$F(x) = \frac{e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{\alpha+(1-\alpha)\,e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}\left[1+\lambda-\frac{\lambda e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{\alpha+(1-\alpha)\,e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}\right], \tag{5}$$

whereas its pdf can be expressed, from (1), (2) and (4) as

$$f(x) = \frac{\alpha\beta\sigma^{\beta}x^{-(\beta+1)}e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{\left[\alpha+(1-\alpha)\,e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{2}}\left[1+\lambda-\frac{2\lambda e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{\alpha+(1-\alpha)\,e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}\right], \tag{6}$$

where $\sigma > 0$ is a scale parameter, α and β are positive shape parameters and $|\lambda| \leq 1$.

A physical interpretation of the cdf of TMOFr is possible if we take a system consisting of two independent components functioning independently at a given time. So, if the two components are connected in parallel, the overall system will have the TMOFr cdf with $\lambda = -1$.

The rf, hrf, reversed hazard rate function (rhrf) and Cumulative hazard rate function (chrf) are, respectively, given by

$$R(x) = \frac{\alpha^2+\left(\alpha-\alpha\lambda-2\alpha^2\right)e^{-\left(\frac{\sigma}{x}\right)^{\beta}}+\left(\alpha^2+\alpha\lambda-\alpha\right)e^{-2\left(\frac{\sigma}{x}\right)^{\beta}}}{\left[\alpha+(1-\alpha)\,e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{2}},$$

$$h(x) = \frac{\alpha\beta\sigma^{\beta}x^{-(\beta+1)}e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{\left[\alpha+(1-\alpha)e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]}\left\{\alpha(1+\lambda)-[\lambda(\alpha+1)+\alpha-1]e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right\}$$
$$\times\left\{\alpha^2+\left[\alpha(1-\lambda-2\alpha)+\left(\alpha^2+\alpha\lambda-\alpha\right)e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right\}^{-1},$$

$$r(x) = \frac{\alpha\beta\sigma^{\beta}x^{-(\beta+1)}\left\{\alpha(1+\lambda)-[\lambda(\alpha+1)+\alpha-1]e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right\}}{\left[\alpha(1+\lambda)-(\alpha\lambda+\alpha-1)e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]\left[\alpha+(1-\alpha)e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]}$$

and

$$H(x) = \ln\left\{\frac{\left[\alpha+(1-\alpha)e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^2}{\alpha^2+(\alpha-\alpha\lambda-2\alpha^2)e^{-\left(\frac{\sigma}{x}\right)^{\beta}}+(\alpha^2+\alpha\lambda-\alpha)e^{-2\left(\frac{\sigma}{x}\right)^{\beta}}}\right\}.$$

Some of the plots of the pdf and hrf of TMOFr for different values of the parameters α, β, σ and λ are displayed in Figures 1 and 2.

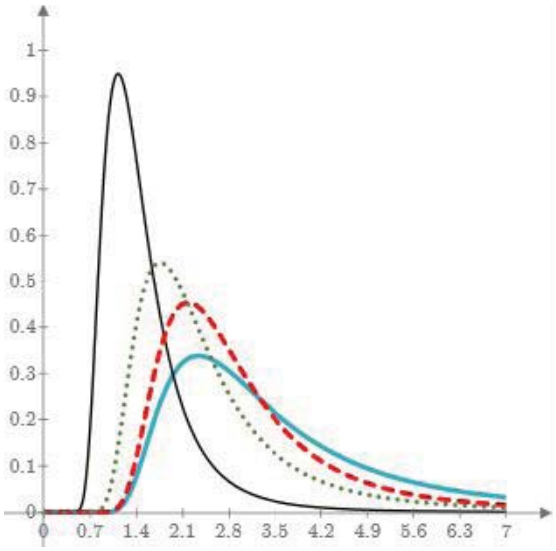


Figure 1. The pdf of TMOFr: (a) For $\alpha = 0.5 : \beta = \lambda = 0.5$ and $\sigma = 2$ (thick line), $\beta = 0.9, \sigma = 0.6$ and $\lambda = -0.2$(black line), $\beta = 0.4, \sigma = 5$ and $\lambda = 1$(dashed line) and $\beta = 0.3, \sigma = 1.5$ and $\lambda = 0.5$.

The TMOFr distribution shows flexible properties as it contains some well known distributions as special cases such as MOFr, transmuted Fréchet (TFr), transmuted inverse exponential (TIE), transmuted inverse Rayleigh (TIR), inverse exponential (IE) and inverse Rayleigh (IR) distributions among others. The flexibility of the TMOFr is explained in Table 1 where it has eleven sub-models when their parameters are carefully chosen.

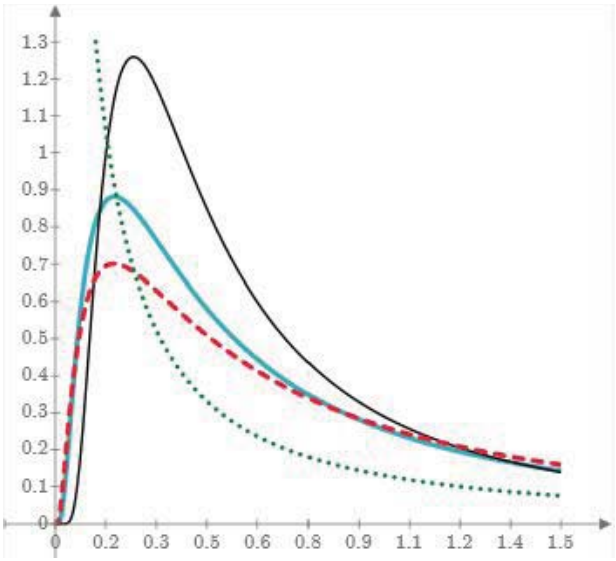


Figure 2. The pdf of TMOFr: (dotted line) (b) For $\alpha = 1.5$ and $\beta = 2.5$: $\sigma = 2.5$ and $\lambda = 0.1$ (thick line), $\sigma = 1.3$ and $\lambda = 0.9$(black line), $\sigma = 2.5$ and $\lambda = 0.7$ (dashed line) and $\sigma = 2$ and $\lambda = 0.6$ (dotted line).

Table 1. Sub-models of the TMOFr

α	β	σ	λ	Reduced Model	Author
α	1	σ	λ	TMOIE	New
α	2	σ	λ	TMOIR	New
α	β	σ	0	MOFr	Krishna et al. (2013)
α	1	σ	0	MOIE	–
α	2	σ	0	MOIR	–
1	β	σ	λ	TFr	Mahmoud and Mandouh (2013)
1	1	σ	λ	TIE	Oguntunde and Adejumo (2015)
1	2	σ	λ	TIR	Ahmad et al. (2014)
1	β	σ	0	Fr	Fréchet (1924)
1	1	σ	0	IE	Keller and Kamath (1982)
1	2	σ	0	IR	Trayer (1964)

3. Mixture Representation

The pdf in (6) can be expressed as

$$f(x)\frac{(1+\lambda)\,\alpha\beta\sigma^{\beta}x^{-(\beta+1)}e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{\left[\alpha+(1-\alpha)\,e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{2}}-\frac{2\lambda\alpha\beta\sigma^{\beta}x^{-(\beta+1)}e^{-2\left(\frac{\sigma}{x}\right)^{\beta}}}{\left[\alpha+(1-\alpha)\,e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{3}}.$$

Expansions for the density of TMOFr can be derived using the series expansion

$$(1-z)^{-k}=\sum_{j=0}^{\infty}\frac{\Gamma(k+j)}{j!\Gamma(k)}z^{j},\ \ |z|<1,\ k>0.$$

Applying the above series expansion, the pdf of the TMOFr can be expressed in the mixture form

$$f(x) = \sum_{k=0}^{\infty} \left[\upsilon_k h_{\beta,\sigma(k+1)^{1/\beta}}(x) - \omega_k h_{\beta,\sigma(k+2)^{1/\beta}}(x) \right], \tag{7}$$

where $\upsilon_k = \frac{(1+\lambda)}{\alpha}\left(1-\frac{1}{\alpha}\right)^k$, $\omega_k = \frac{\nu(k+1)}{\alpha^2}\left(1-\frac{1}{\alpha}\right)^k$ and $h_{\beta,\sigma(\delta)^{1/\beta}}(x)$ is the Fréchet (Fr) density with shape parameter β and scale parameter $\sigma(\delta)^{1/\beta}$.

Since the density function TMOFr is expressed as a mixture of Fr densities, one may obtain some of its mathematical properties directly from the properties of the Fr distribution.

By integrating (7), we obtain

$$F(x) = \sum_{k=0}^{\infty} \left[\upsilon_k H_{\beta,\sigma(k+1)^{1/\beta}}(x) - \omega_k H_{\beta,\sigma(k+2)^{1/\beta}}(x) \right],$$

where $H_{\beta,\sigma(\delta)^{1/\beta}}(x)$ is the cdf of Fr distribution with shape parameter β and scale parameter $\sigma(\delta)^{1/\beta}$.

4. Mathematical Properties

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

4.1 Moments

Henceforth, let Z be a random variable having the Fr distribution with scale $\sigma > 0$ and shape $\beta > 0$. Then, the pdf of Z is given by

$$g(z;\beta,\sigma) = \beta\sigma^{\beta} z^{-(\beta+1)} e^{-\left(\frac{\sigma}{z}\right)^{\beta}},\ z > 0.$$

For $r < \beta$, the rth ordinary and incomplete moments of Z are given by

$$\mu'_{r,Z} = \sigma^r \Gamma\left(1-\frac{r}{\beta}\right) \quad \text{and} \quad \varphi_{r,Z}(t) = \sigma^r \gamma\left(1-\frac{r}{\beta}, (\sigma/t)^{\beta}\right),$$

respectively, where $\gamma(s,t) = \int_0^t x^{s-1} e^{-x} dx$ is the lower incomplete gamma function.

Then, the rth moment of X, say μ'_r, can be expressed as

$$\mu'_r = E(X^r) = \sum_{k=0}^{\infty} \sigma^r \left[\upsilon_k\ (k+1)^{r/\beta} - \omega_k\ (k+2)^{r/\beta} \right] \Gamma\left(1-\frac{r}{\beta}\right). \tag{8}$$

Using the relation between the central and non-central moments, we obtain the nth central moment of X, say μ_n, as follows

$$\mu_n = \sum_{r=0}^{n}\sum_{k=0}^{\infty} \sigma^r (-\mu'_1)^{n-r} \left[\upsilon_k\ (k+1)^{r/\beta} - \omega_k\ (k+2)^{r/\beta} \right] \binom{n}{r} \Gamma\left(1-\frac{r}{\beta}\right).$$

The skewness and kurtosis measures can be determined from the central moments using established results.

4.2 Incomplete Moments

The main application of the first incomplete moment refers to the Bonferroni and Lorenz curves. These curves are very useful in economics, reliability, demography, insurance and medicine. The answers to many important questions in economics require more than just knowing the mean of the distribution, but its shape as well. This is obvious not only in the study of econometrics but in other areas as well. The sth incomplete moments, say $\varphi_s(t)$, is given by

$$\varphi_s(t) = \int_0^t x^s f(x)\, dx,$$

Using equation (7), we can write

$$\varphi_s(t) = \sum_{k=0}^{\infty} \left[\upsilon_k \int_0^t x^s h_{\beta,\sigma(k+1)^{1/\beta}}(x) - \omega_k \int_0^t x^s h_{\beta,\sigma(k+2)^{1/\beta}}(x) \right],$$

and then using the lower incomplete gamma function, we obtain (for $s < \alpha$)

$$\varphi_s(t) = \sum_{k=0}^{\infty} \upsilon_k \sigma^s (k+1)^{s/\beta} \gamma\left(1 - \frac{s}{\beta}, (k+1)\left(\frac{\sigma}{t}\right)^{\beta}\right) - \sum_{k=0}^{\infty} \omega_k \sigma^s (k+2)^{s/\beta} \gamma\left(1 - \frac{s}{\beta}, (k+2)\left(\frac{\sigma}{t}\right)^{\beta}\right),$$

where $\gamma(a, z)$ is, the lower incomplete gamma function, defined in subsection 4.1.

The first incomplete moment of the TMOFr distribution can be obtained by setting $s = 1$ in the last equation.

Another application of the first incomplete moment is related to mean residual life and mean waiting time given by $m_1(t) = [1 - \varphi_1(t)]/R(t) - t$ and $M_1(t) = t - [\varphi_1(t)/F(t)]$, respectively.

The amount of scattering in a population is evidently measured, to some extent, by the totality of deviations from the mean and median. The mean deviations about the mean $[\delta_\mu(X) = E(|X - \mu_1^{'}|)]$ and about the median $[\delta_\mu(X) = E(|X - M|)]$ of X can be, used as measures of spread in a population, expressed by $\delta_\mu(X) = \int_0^{\infty} |X - \mu_1^{'}| f(x)\, dx = 2\mu_1^{'} F(\mu_1^{'}) - 2\varphi_1(\mu_1^{'})$ and $\delta_M(X) = \int_0^{\infty} |X - M| f(x)\, dx = \mu_1^{'} - 2\varphi_1(M)$, respectively, where $\mu_1^{'} = E(X)$ comes from (8), $F(\mu_1^{'})$ is simply calculated, $\varphi_1(\mu_1^{'})$ is the first incomplete moments and M is the median of X.

4.3 Residual Life Function

Several functions are defined related to the residual life. The failure rate function, mean residual life function and the left censored mean function, also called vitality function. It is well known that these three functions uniquely determine $F(x)$, see Gupta (1975), Kotz and Shanbhag (1980) and Zoroa et al. (1990).

Moreover, the nth moment of the residual life, say $m_n(t) = E[(X - t)^n \mid X > t]$, $n = 1, 2, \ldots$, uniquely determine $F(x)$ (see Navarro et al., (1998). The nth moment of the residual life of X is given by

$$m_n(t) = \frac{1}{R(t)} \int_t^{\infty} (x - t)^n f(x) dx.$$

Then, we can write (for $r < \beta$)

$$m_n(t) = \frac{1}{R(t)} \sum_{r=0}^{n} \sum_{n=0}^{\infty} (-t)^{n-r} \upsilon_k \sigma^r (k+1)^{\frac{r}{\beta}} \binom{n}{r} \left\{1 - \gamma\left(1 - \frac{r}{\beta}, (k+1)\left(\frac{\sigma}{t}\right)^{\beta}\right)\right\}$$
$$- \frac{1}{R(t)} \sum_{r=0}^{n} \sum_{n=0}^{\infty} (-t)^{n-r} \omega_k \sigma^r (k+2)^{\frac{r}{\beta}} \binom{n}{r} \left\{1 - \gamma\left(1 - \frac{r}{\beta}, (k+2)\left(\frac{\sigma}{t}\right)^{\beta}\right)\right\}.$$

Another interesting function is the mean residual life (MRL) function or the life expectation at age x defined by $m_1(x) = E[(X - x) \mid X > x]$, which represents the expected additional life length for a unit which is alive at age x. The MRL of the TMOFr distribution can be obtained by setting $n = 1$ in the last equation. Guess and Proschan (1988) derived an extensive coverage of possible applications of the MRL applications in survival analysis, biomedical sciences, life insurance, maintenance and product quality control, economics, social studies and demography (see, Lai and Xie, 2006).

4.4 Reversed Residual Life Function

The nth moment of the reversed residual life, say $M_n(t) = E[(t - X)^n \mid X \leq t]$ for $t > 0$, $n = 1, 2, \ldots$ uniquely determines $F(x)$ (Navarro et al., 1998). We obtain

$$M_n(t) = \int_0^t (t - x)^n dF(x).$$

Therefore, the nth moment of the reversed residual life of X given that $r < \beta$ becomes

$$M_n(t) = \frac{1}{F(t)} \sum_{r=0}^{n} \sum_{n=0}^{\infty} (-1)^r t^{n-r} \upsilon_k \sigma^r (k+1)^{\frac{r}{\beta}} \binom{n}{r} \gamma\left(1 - \frac{r}{\beta}, (k+1)\left(\frac{\sigma}{t}\right)^{\beta}\right)$$
$$- \frac{1}{F(t)} \sum_{r=0}^{n} \sum_{n=0}^{\infty} (-1)^r t^{n-r} \omega_k \sigma^r (k+2)^{\frac{r}{\beta}} \binom{n}{r} \gamma\left(1 - \frac{r}{\beta}, (k+2)\left(\frac{\sigma}{t}\right)^{\beta}\right).$$

The mean inactivity time (MIT) or mean waiting time (MWT), also called the mean reversed residual life function, is defined by $M_1(t) = E[(t - X) \mid X \leq t]$, and it represents the waiting time elapsed since the failure of an item on condition that this failure had occurred in $(0, x)$. The MRRL of X can be obtained by setting $n = 1$ in the above equation. The properties of the mean inactivity time have been considered by many authors, see e.g., Kayid and Ahmad (2004) and Ahmad et al. (2005).

4.5 Quantile and Generating Functions

The quantile function (qf) of X is the real solution of the equation $F(x_q) = q$. Then by inverting (5), we obtain

$$x_q = \sigma \left[\ln \left(\frac{b + \alpha \sqrt{1 + \lambda(\lambda - 4q + 2)}}{2\alpha^2 q} \right) \right]^{-1/\beta}, 0 < q < 1,$$

where $b = 2\alpha q\,(\alpha - 1) + \alpha\,(\lambda + 1)$.

Simulating a TMOFr random variable is straightforward. If U is a uniform variate on the unit interval $(0, 1)$, then the random variable $X = x_q$ follows (6).

First, we provide the moment generating function (mgf) of the Fr model as discussed by Afify et al. (2015). We can write the mgf of Z as

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

By expanding the first exponential and determining the integral, we obtain

$$M(t; \beta, \sigma) = \sum_{m=0}^{\infty} \frac{\sigma^m\, t^m}{m!}\, \Gamma\left(\frac{\beta - m}{\beta} \right).$$

Consider the Wright generalized hypergeometric function (Kilbas et al., 2006) defined by

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

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

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

Combining the last expression and (7), the mgf of X can be expressed as

$$M_X(t) = \sum_{k=0}^{\infty} \upsilon_k\, {}_1\Psi_0 \left[\begin{array}{c} \left(1, -\beta^{-1}\right) \\ - \end{array} ; \sigma\,(k+1)^{1/\beta}\, t \right] - \sum_{k=0}^{\infty} \omega_k\, {}_1\Psi_0 \left[\begin{array}{c} \left(1, -\beta^{-1}\right) \\ - \end{array} ; \sigma\,(k+2)^{1/\beta}\, t \right].$$

4.6 Rényi and q-Entropies

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

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

Then, using (6), we can write

$$f(x)^{\gamma} = \frac{\left(\alpha\beta\sigma^{\beta}\right)^{\gamma} x^{-\gamma(\beta+1)} e^{-\gamma\left(\frac{\sigma}{x}\right)^{\beta}}}{\left[\alpha + (1 - \alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{2\gamma}} \underbrace{\left[1 + \lambda - 2\lambda \frac{e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{\alpha + (1 - \alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}\right]^{\gamma}}_{A}.$$

Applying the generalized binomial expansion to the quantity A, we obtain

$$f(x)^{\gamma} = \left[\alpha\beta\sigma^{\beta}\,(1 + \lambda)\right]^{\gamma} x^{-\gamma(\beta+1)} \sum_{j=0}^{\infty} \frac{(-1)^j\, \Gamma(\gamma + 1)\, d^j}{j!\, \Gamma(\gamma - j + 1)} \times e^{-(j+\gamma)\left(\frac{\sigma}{x}\right)^{\beta}} \underbrace{\left[\alpha + (1 - \alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-(2\gamma+j)}}_{B},$$

where $d = 2\lambda/(1+\lambda)$.

Applying the series expansion defined in Section 3 to B, we can write

$$f(x)^{\gamma} = \left(\beta\sigma^{\beta}\right)^{\gamma} \sum_{j,k=0}^{\infty} b_{j,k}\, x^{-\gamma(\beta+1)} e^{-(k+j+\gamma)\left(\frac{\sigma}{x}\right)^{\beta}},$$

where

$$b_{j,k} = \frac{(-1)^j \Gamma(2\gamma + j + k)\,\Gamma(\gamma+1)\, d^j (1+\lambda)^{\gamma}}{j!k!\Gamma(2\gamma + j)\,\Gamma(\gamma - j + 1)\,\alpha^{\gamma+j}} \left(1 - \frac{1}{\alpha}\right)^k.$$

Then the Rényi entropy of X is given by

$$I_{\gamma}(X) = \frac{1}{(\gamma-1)} \log\left[\left(\beta\sigma^{\beta}\right)^{\gamma} \sum_{j,k=0}^{\infty} b_{j,k} \int_{-\infty}^{\infty} x^{-\gamma(\beta+1)} e^{-(k+j+\gamma)\left(\frac{\sigma}{x}\right)^{\beta}} dx\right],$$

and by making the substitution $u = (\sigma/x)^{\beta}$, for $\gamma(1+\beta) > 1$, we have that

$$I_{\gamma}(X) = \frac{1}{(\gamma-1)} \log\left[\left(\frac{\beta}{\sigma}\right)^{\gamma-1} \sum_{j,k=0}^{\infty} b_{j,k}\,(k+j+\gamma)^{-s}\,\Gamma(s)\right],$$

where $s = [\gamma(1+\beta) - 1]/\beta$.

The q-entropy (for $q > 0$ and $q \neq 1$), say $H_q(X)$, is given by

$$H_q(X) = \frac{1}{(q-1)} \log\left\{1 - \int_{-\infty}^{\infty} f(x)^q\, dx\right\},$$

and then

$$H_q(X) = \frac{1}{(q-1)} \log\left\{1 - \left[\left(\frac{\beta}{\sigma}\right)^{q-1} \sum_{j,k=0}^{\infty} b^{\star}_{j,k}\,(k+j+q)^{-s^{\star}}\,\Gamma\left(s^{\star}\right)\right]\right\},$$

where $s^{\star} = [q(1+\beta) - 1]/\beta$ and

$$b^{\star}_{j,k} = \frac{(-1)^j \Gamma(2q + j + k)\,\Gamma(q+1)\, d^j (1+\lambda)^{q}}{j!k!\Gamma(2q + j)\,\Gamma(q - j + 1)\,\alpha^{q+j}} \left(1 - \frac{1}{\alpha}\right)^k.$$

5. Order Statistics

The order statistics and their moments have great importance in many statistical problems and they have many applications in reliability analysis and life testing. Let $X_1, \ldots, X_n$ be a random sample of size n from the TMOFr(α, β, σ, λ) with cdf (5) and pdf (6), respectively. Let $X_{1:n}, \ldots, X_{n:n}$ be the corresponding order statistics. Then, the pdf of rth order statistic, say $X_{r:n}$, $1 \leq r \leq n$, denoted by $f_{r:n}(x)$, can be expressed as

$$f_{r:n}(x) = C_{r:n} \frac{\alpha\beta\sigma^{\beta} x^{-(\beta+1)} e^{-\left(\frac{\sigma}{x}\right)^{\beta}} \left[\ell_1 e^{-\left(\frac{\sigma}{x}\right)^{\beta}} - \ell_3 e^{-2\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{r-1}}{\left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{3} \left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{2(r-1)}} \times \frac{\left[\ell_1 - \ell_2 e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]\left[\alpha^2 + \ell_4 e^{-\left(\frac{\sigma}{x}\right)^{\beta}} + \ell_5 e^{-2\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{n-r}}{\left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{2(n-r)}},$$

where $C_{r:n} = \frac{n!}{(r-1)!(n-r)!}$, $\ell_1 = \alpha(1+\lambda)$, $\ell_2 = \alpha\lambda + \alpha + \lambda - 1$, $\ell_3 = \alpha\lambda + \alpha - 1$, $\ell_4 = \alpha - \alpha\lambda - 2\alpha^2$ and $\ell_5 = \alpha^2 + \alpha\lambda - \alpha$.

The pdf $f_{r:n}(x)$, can also be expressed as

$$f_{r:n}(x) = \frac{f(x)}{\mathrm{B}(r, n-r+1)} \sum_{s=0}^{n-1} (-1)^s \binom{n-1}{s} F^{s+r-1}(x). \tag{9}$$

Further, we can write

$$F^{s+r-1}(x) = \sum_{i=0}^{\infty} (-\lambda)^i \binom{s+r-1}{i} \frac{(1+\lambda)^{s+r-i-1}\, e^{-(s+r+i-1)\left(\frac{\sigma}{x}\right)^{\beta}}}{\left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{s+r+i-1}}.$$

Using equation (6) and the last equation and, after some simplification, we can write

$$f(x)F^{s+r-1}(x) = (1+\lambda)\sum_{i=0}^{\infty}\frac{d_i\alpha\beta\sigma^{\beta}x^{-(\beta+1)}e^{-(s+r+i)\left(\frac{\sigma}{x}\right)^{\beta}}}{\left[\alpha+(1-\alpha)\,e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{s+r+i+1}} - 2\lambda\sum_{i=0}^{\infty}\frac{d_i\alpha\beta\sigma^{\beta}x^{-(\beta+1)}e^{-(s+r+i+1)\left(\frac{\sigma}{x}\right)^{\beta}}}{\left[\alpha+(1-\alpha)\,e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{s+r+i+2}}, \tag{10}$$

where $d_i = (-\lambda)^i\binom{s+r-1}{i}(1+\lambda)^{s+r-i-1}$.

By inserting (10) in equation (9) and, after some simplification, we obtain

$$f_{r:n}(x) = \sum_{s=0}^{n-1}\sum_{i,j=0}^{\infty} a_{i,j}\,h_{s+r+i+j}(x) - \sum_{s=0}^{n-1}\sum_{i,j=0}^{\infty} b_{i,j}\,h_{s+r+i+j+1}(x), \tag{11}$$

where

$$a_{i,j} = \frac{(-1)^s\,\Gamma(s+r+i+j+1)\,(1+\lambda)\,d_i\left(1-\frac{1}{\alpha}\right)^j}{j!\mathrm{B}(r,n-r+1)\,\Gamma(s+r+i+1)\,(s+r+i+j)\,\alpha^{s+r+i}}\binom{n-1}{s},$$

$$b_{i,j} = \frac{(-1)^s\,\Gamma(s+r+i+j+2)\,2\lambda d_i\left(1-\frac{1}{\alpha}\right)^j}{j!\mathrm{B}(r,n-r+1)\,\Gamma(s+r+i+2)\,(s+r+i+j+1)\,\alpha^{s+r+i+1}}\binom{n-1}{s}$$

and $h_\eta(x)$ denotes to the Fr density with shape parameter β and scale parameter $\sigma(\eta)^{1/\beta}$.

Equation (11) reveals that the pdf of the TMOFr order statistics is a mixture of Fr densities. So, some of their mathematical properties can also be obtained from those of the Fr distribution. For example, the qth moment of $X_{r:n}$ can be expressed as

$$E\left(X_{r:n}^q\right) = \sum_{s=0}^{n-1}\sum_{i,j=0}^{\infty} a_{i,j}\,E\left(Y_{s+r+i+j}^q\right) - \sum_{s=0}^{n-1}\sum_{i,j=0}^{\infty} b_{i,j}\,E\left(Y_{s+r+i+j+1}^q\right), \tag{12}$$

where $Y_{s+r+i+j}^q \sim \mathrm{Fr}(\sigma(s+r+i+j)^{1/\beta},\beta)$.

The L-moments are analogous to the ordinary moments but can be estimated by linear combinations of order statistics. Based upon the moments in equation (12), we can derive explicit expressions for the L-moments of X as infinite weighted linear combinations of the means of suitable TMOFr distribution. They are linear functions of expected order statistics defined by

$$\lambda_r = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}E(X_{r-k:r}),\ r\geq 1.$$

The first four L-moments are given by

$$\lambda_1 = E(X_{1:1}),\quad \lambda_2 = \frac{1}{2}E(X_{2:2}-X_{1:2}),\lambda_3 = \frac{1}{3}E(X_{3:3}-2X_{2:3}+X_{1:3})$$

and

$$\lambda_4 = \frac{1}{4}E(X_{4:4}-3X_{3:4}+3X_{2:4}-X_{1:4}).$$

6. Characterizations

The problem of characterizing a distribution is an important problem in various fields which has recently attracted the attention of many researchers. These characterizations have been established in many different directions. This section deals with various characterizations of TMOFr distribution. These characterizations are based on: (*i*) a simple relationship between two truncated moments; (*ii*) the hazard function; (*iii*) a single function of the random variable. It should be mentioned that for characterization (*i*) the cdf need no have a closed form. We believe, due to the nature of the cdf of TMOFr, there may not be other possible characterizations than the ones presented in this section.

6.1 Characterizations Based on Two Truncated Moments

In this subsection we present characterizations of TMOFr distribution in terms of a simple relationship between two truncated moments. Our first characterization result borrows from a theorem due to Glanzel (1987) see Theorem A of Appendix A. We refer the interested reader to Glanzel (1987) for a proof of Theorem A. Note that the result holds also when the interval H is not closed. Moreover, as mentioned above, it could be also applied when the cdf F does not have a closed form. As shown in Afify et al. (2015), this characterization is stable in the sense of weak convergence.

Proposition 6.1. Let $X : \Omega \to (0, \infty)$ be a continuous random variable and let $h(x) = \left[1 + \lambda - \frac{\lambda e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{\alpha+(1-\alpha)e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}\right]^{1-b}$ and $g(x) = h(x)\left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-1}$ for $x > 0$. The random variable X belongs to TMOFr family (6) if and only if the function η defined in Theorem A has the form

$$\eta(x) = \frac{1}{2}\left\{1 + \left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-1}\right\}, \quad x > 0. \tag{13}$$

Proof. Let X be a random variable with density (6), then

$$(1 - F(x))\, E\left[h(x) \mid X \geq x\right] = \frac{1}{1-\alpha}\left\{\left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-1} - 1\right\},\ x > 0,$$

and

$$(1 - F(x))\, E\left[g(x) \mid X \geq x\right] = \frac{1}{2(1-\alpha)}\left\{\left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-2} - 1\right\},\ x > 0,$$

and finally

$$\eta(x)\, h(x) - g(x) = \frac{1}{2} h(x)\left\{1 - \left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-1}\right\} > 0 \text{ for } x > 0.$$

Conversely, if η is given as above, then

$$s'(x) = \frac{\eta'(x)\, h(x)}{\eta(x)\, h(x) - g(x)} = \frac{-(1-\alpha)\beta\sigma^{\beta} e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-2}}{x^{\beta+1}\left\{1 - \left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-1}\right\}},\ x > 0,$$

and hence

$$s(x) = -\ln\left\{\left\{1 - \left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-1}\right\}\right\},\ x > 0.$$

Now, in view of Theorem A, X has density (6).

Corollary 6.1. Let $X : \Omega \to (0, \infty)$ be a continuous random variable and let $h(x)$ be as in Proposition 6.1. The pdf of X is (6) if and only if there exist functions g and η defined in Theorem A satisfying the differential equation

$$\frac{\eta'(x)\, h(x)}{\eta(x)\, h(x) - g(x)} = \frac{-(1-\alpha)\beta\sigma^{\beta} e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-2}}{x^{\beta+1}\left\{1 - \left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-1}\right\}},\ x > 0. \tag{14}$$

The general solution of the differential equation in Corollary 6.1 is

$$\eta(x) = \left\{1 - \left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-1}\right\}^{-1}\left\{\int \frac{(1-\alpha)\beta\sigma^{\beta} x^{-(\beta+1)} e^{-\left(\frac{\sigma}{x}\right)^{\beta}}}{h(x)\left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{2}} g(x)\, dx + D\right\}$$

where D is a constant. Note that a set of functions satisfying the differential equation (14) is given in Proposition 6.1 with $D = \frac{1}{2}$. However, it should be also noted that there are other triplets (h, g, η) satisfying the conditions of Theorem A.

6.2 Characterization Based on Hazard Function

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

$$\frac{f'(x)}{f(x)} = \frac{h'_F(x)}{h_F(x)} - h_F(x). \tag{15}$$

For many univariate continuous distributions, this is the only characterization available in terms of the hazard function. The following characterization establishes a non-trivial characterization for TMOFr distribution in terms of the hazard function when $\alpha = 1$, which is not of the trivial form given in (15). We assume, without loss of generality, that $\sigma = 1$ in the following Proposition.

Proposition 6.2. Let $X : \Omega \to (0, \infty)$ be a continuous random variable. Then for $\alpha = 1$, the pdf of X is (6) if and only if its hazard function $h_F(x)$ satisfies the differential equation

$$h'_F(x) + (\beta + 1) x^{-1} h_F(x) = \beta x^{-(\beta+1)} \frac{d}{dx} \left\{ \frac{e^{-\left(\frac{\sigma}{x}\right)^\beta} \left[1 + \lambda - 2\lambda e^{-\left(\frac{\sigma}{x}\right)^\beta}\right]}{1 + \left[\lambda e^{-\left(\frac{\sigma}{x}\right)^\beta} - (1 + \lambda)\right] e^{-\left(\frac{\sigma}{x}\right)^\beta}} \right\}, \tag{16}$$

with the boundary condition $\lim_{x\to\infty} h_F(x) = 0$.

Proof. If X has pdf (6), then clearly (16) holds. Now, if (16) holds, then

$$\frac{d}{dx}\left\{x^{\beta+1} h_F(x)\right\} = \frac{d}{dx}\left\{ \frac{\beta e^{-\left(\frac{\sigma}{x}\right)^\beta} \left[1 + \lambda - 2\lambda e^{-\left(\frac{\sigma}{x}\right)^\beta}\right]}{1 + \left[\lambda e^{-\left(\frac{\sigma}{x}\right)^\beta} - (1 + \lambda)\right] e^{-\left(\frac{\sigma}{x}\right)^\beta}} \right\},$$

or, equivalently,

$$h_F(x) = \frac{\beta x^{-(\beta+1)} e^{-\left(\frac{\sigma}{x}\right)^\beta} \left[1 + \lambda - 2\lambda e^{-\left(\frac{\sigma}{x}\right)^\beta}\right]}{1 + \left[\lambda e^{-\left(\frac{\sigma}{x}\right)^\beta} - (1 + \lambda)\right] e^{-\left(\frac{\sigma}{x}\right)^\beta}},$$

which is the hazard function of the TMOFr distribution.

6.3 Characterizations Based on a Single Function of the Random Variable

In this subsection we present a characterization result in terms of a function of the random variable X.

Proposition 6.3. Let $X : \Omega \to (a, b)$ be a continuous random variable with cdf F and corresponding pdf f. Let $\psi(x)$ be a differentiable function greater than 1 on (a, b) such that $\lim_{x\to a^+} \psi(x) = 1$ and $\lim_{x\to b^-} \psi(x) = 1 + c$. Then, for $0 < c < 1$,

$$E\left[\psi(X) \mid X \le x\right] = c + (1 - c)\psi(x), \tag{17}$$

if and only if

$$F(x) = \left(\frac{\psi(x) - 1}{c}\right)^{\frac{1-c}{c}}. \tag{18}$$

Proof. If (17) holds, then

$$\int_a^x \psi(u) f(u)\, du = \{c + (1 - c)\psi(x)\} F(x).$$

Differentiating both sides of the above equation with respect to x and rearranging the terms, we arrive at

$$\frac{f(x)}{F(x)} = \frac{1-c}{c}\left(\frac{\psi'(x)}{\psi(x)-1}\right). \tag{19}$$

Integrating both sides of (19) from x to b and using the condition $\lim_{x\to b^-}\psi(x) = 1+c$, we arrive at (18). Conversely, if (18) holds, then $\psi(x) = 1 + c\,(F(x))^{\frac{c}{1-c}}$ and

$$\begin{aligned} E\left[\psi(X) \mid X \le x\right] &= \frac{\int_a^x \left\{1 + c\,(F(u))^{\frac{c}{1-c}}\right\} f(u)\,du}{F(x)} \\ &= \frac{F(x) + c\,(1-c)\,(F(x))^{\frac{1}{1-c}}}{F(x)} \\ &= 1 + c\,(1-c)\,(F(x))^{\frac{c}{1-c}} \\ &= 1 + (1-c)\,(\psi(x)-1) \\ &= c + (1-c)\,\psi(x), \end{aligned}$$

which is (17).

Remark 6.1. Taking, e.g., $(a,b) = (0,\infty)$ and

$$\psi(x) = 1 + c\left\{\frac{e^{-(\frac{\sigma}{x})\beta}}{\alpha + (1-\alpha)\,e^{-(\frac{\sigma}{x})\beta}}\left[1 + \lambda - \frac{\lambda e^{-(\frac{\sigma}{x})\beta}}{\alpha + (1-\alpha)\,e^{-(\frac{\sigma}{x})\beta}}\right]\right\}^{\frac{c}{1-c}}.$$

Proposition 6.3 gives a characterization of TMOFr distribution.

7. Maximum Likelihood Estimation

Several approaches for parameter estimation were proposed in the literature but the maximum likelihood method is the most commonly employed. The MLEs enjoy desirable properties and can be used when constructing confidence intervals and regions and also in test statistics. The normal approximation for these estimators in large sample distribution theory is easily handled either analytically or numerically. In this section, we consider the estimation of the parameters of the TMOFr$(\alpha,\beta,\sigma,\lambda,x)$ distribution by maximum likelihood. Consider the random sample $x_1,\ldots,x_n$ of size n from this distribution. The log-likelihood function for the parameter vector $\phi = (\alpha,\beta,\sigma,\lambda)^{\intercal}$, say (ϕ), is given by

$$\begin{aligned} (\phi) &= n\left(\log\alpha + \ln\beta - \log\sigma\right) - \sum_{i=1}^{n}\left(\frac{\sigma}{x_i}\right)^{\beta} + (\beta+1)\sum_{i=1}^{n}\log\left(\frac{\sigma}{x_i}\right) \\ &\quad -3\sum_{i=1}^{n}\log\left[\alpha + (1-\alpha)\,e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}\right] + \sum_{i=1}^{n}\log\left[\alpha(\lambda+1) - pe^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}\right], \end{aligned}$$

where $p = \lambda(\alpha+1)+\alpha-1$. The above equation can be maximized either directly by using the `MATH-CAD` program, `R` (`optim` function), `SAS` (`PROC NLMIXED`) or by solving the nonlinear equations obtained by differentiating the log-likelihood. Therefore, the corresponding score function, say $\mathbf{U}(\phi) = \frac{\partial(\phi)}{\partial\phi}$, is given by $\mathbf{U}(\phi) = \left(\frac{\partial(\phi)}{\partial\alpha}, \frac{\partial(\phi)}{\partial\beta}, \frac{\partial(\phi)}{\partial\sigma}, \frac{\partial(\phi)}{\partial\lambda}\right)^{\intercal}$. Then,

$$\frac{\partial(\phi)}{\partial\alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n}\frac{3 - 3e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}{\alpha + (1-\alpha)\,e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}} + \sum_{i=1}^{n}\frac{1 + \lambda - (1+\lambda)\,e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}{\alpha(\lambda+1) - pe^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}},$$

$$\frac{\partial(\phi)}{\partial\beta} = \sum_{i=1}^{n}\left[1 - \left(\frac{\sigma}{x_i}\right)^{\beta}\right]\log\left(\frac{\sigma}{x_i}\right) + \sum_{i=1}^{n}\frac{p\left(\frac{\sigma}{x_i}\right)^{\beta}e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}{\alpha(\lambda+1) - pe^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}\log\left(\frac{\sigma}{x_i}\right) + \frac{n}{\beta} + \sum_{i=1}^{n}\frac{3(1-\alpha)\left(\frac{\sigma}{x_i}\right)^{\beta}e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}{\alpha + (1-\alpha)\,e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}\log\left(\frac{\sigma}{x_i}\right),$$

$$\frac{\partial(\phi)}{\partial\sigma} = \frac{1}{\sigma}\left[n - \beta\sum_{i=1}^{n}\left(\frac{\sigma}{x_i}\right)^{\beta}\right] - \frac{\beta}{\sigma}\sum_{i=1}^{n}\frac{p\left(\frac{\sigma}{x_i}\right)^{\beta} e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}{\alpha(\lambda+1) - pe^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}} + \frac{\beta}{\sigma}\sum_{i=1}^{n}\frac{3(1-\alpha)\, e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}{\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}\left(\frac{\sigma}{x_i}\right)^{\beta}$$

and

$$\frac{\partial(\phi)}{\partial\lambda} = \sum_{i=1}^{n}\frac{\alpha - (1+\alpha)\, e^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}{\alpha(\lambda+1) - pe^{-\left(\frac{\sigma}{x_i}\right)^{\beta}}}.$$

We can obtain the estimates of the unknown parameters by setting the score vector to zero, $\mathbf{U}(\widehat{\phi}) = 0$. Solving these equations simultaneously gives the MLEs $\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}$ and $\widehat{\lambda}$. If they can not be solved analytically and statistical software can be used to solve them numerically by means of iterative techniques such as the Newton-Raphson algorithm. For the TMOFr distribution all the second order derivatives exist.

For interval estimation of the model parameters, we require the 4×4 observed information matrix $J(\phi) = \{J_{rs}\}$ for $r, s = \alpha, \beta, \sigma, \lambda$. Under standard regularity conditions, the multivariate normal $N_4(0, J(\widehat{\phi})^{-1})$ distribution can be used to construct approximate confidence intervals for the model parameters. Here, $J(\widehat{\phi})$ is the total observed information matrix evaluated at $\widehat{\phi}$. Therefore, approximate $100(1-\varphi)\%$ confidence intervals for α, β, σ and λ can be determined as:

$\widehat{\alpha} \pm z_{\frac{\varphi}{2}}\sqrt{\widehat{J}_{\alpha\alpha}}$, $\widehat{\beta} \pm z_{\frac{\varphi}{2}}\sqrt{\widehat{J}_{\beta\beta}}$, $\widehat{\sigma} \pm z_{\frac{\varphi}{2}}\sqrt{\widehat{J}_{\sigma\sigma}}$ and $\widehat{\lambda} \pm z_{\frac{\varphi}{2}}\sqrt{\widehat{J}_{\lambda\lambda}}$, where $z_{\frac{\varphi}{2}}$ is the upper φth percentile of the standard normal distribution.

8. Applications

In this section, We provide two applications to two real data sets to prove the flexibility of the TMOFr model. We compare the fit of the TMOFr with competitve models namely: MOFr, BFr, GEFr, TFr and Fr distributions. The pdfs of these distributions are, respectively, given by (for $x > 0$):

MOFr: $f(x) = \alpha\beta\sigma^{\beta}\, x^{-(\beta+1)}\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\left[\alpha + (1-\alpha)\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{-2}$;

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

GEFr: $f(x) = \frac{a\beta\sigma^{\beta}}{\Gamma(b)}\, x^{-(\beta+1)}\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\left[1 - e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{a-1}\left\{-\log\left[1 - e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]^{a}\right\}^{b-1}$;

TFr: $f(x) = \beta\sigma^{\beta}x^{-(\beta+1)}\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\left[1 + \lambda - 2\lambda\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}\right]$;

Fr: $f(x) = \beta\sigma^{\beta}x^{-(\beta+1)}\, e^{-\left(\frac{\sigma}{x}\right)^{\beta}}$.

The parameters of the above densities are all positive real numbers except for the TFr distribution for which $|\lambda| \leq 1$.

The first data set refers to breaking stress of carbon fibres (in Gba) (Nichols and Padgett, 2006) and consists of 100 observations: 3.70, 2.74, 2.73, 2.50, 3.60, 3.11, 3.27, 2.87, 1.47, 3.11, 4.42, 2.41, 3.19, 3.22, 1.69, 3.28, 3.09, 1.87, 3.15, 4.90, 3.75, 2.43, 2.95, 2.97, 3.39, 2.96, 2.53, 2.67, 2.93, 3.22, 3.39, 2.81, 4.20, 3.33, 2.55, 3.31, 3.31, 2.85, 2.56, 3.56, 3.15, 2.35, 2.55, 2.59, 2.38, 2.81, 2.77, 2.17, 2.83, 1.92, 1.41, 3.68, 2.97, 1.36, 0.98, 2.76, 4.91, 3.68, 1.84, 1.59, 3.19, 1.57, 0.81, 5.56, 1.73, 1.59, 2.00, 1.22, 1.12, 1.71, 2.17, 1.17, 5.08, 2.48, 1.18, 3.51, 2.17, 1.69, 1.25, 4.38, 1.84, 0.39, 3.68, 2.48, 0.85, 1.61, 2.79, 4.70, 2.03, 1.80, 1.57, 1.08, 2.03, 1.61, 2.12, 1.89, 2.88, 2.82, 2.05, 3.65. The second data set is obtained from Smith and Naylor (1987). The data are the strengths of 1.5 cm glass fibres, measured at the National Physical Laboratory, England. Unfortunately, the units of measurement are not given in the paper. The data set consisting of 63 observations are: 0.55, 0.93,1.25, 1.36, 1.49, 1.52, 1.58, 1.61, 1.64, 1.68, 1.73, 1.81, 2 , 0.74, 1.04, 1.27, 1.39, 1.49, 1.53,1.59, 1. 61, 1.66, 1.68, 1.76, 1.82, 2.01, 0.77, 1.11, 1.28, 1.42, 1.5, 1.54, 1.6, 1.62, 1.66, 1.69,1.76, 1.84, 2.24, 0.81, 1.13, 1.29, 1.48, 1.5, 1.55, 1.61, 1.62, 1.66, 1.7, 1.77, 1.84, 0.84, 1.24, 1.3, 1.48, 1.51, 1.55, 1.61, 1.63, 1.67, 1.7, 1.78, 1.89.

In order to compare the distributions, we consider the measures of goodness-of-fit including the Akaike information criterion (AIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion ($HQIC$) and consistent

Akaike information criterion ($CAIC$). These measures of goodness-of-fit evaluated at the MLEs$\widehat{\ }$, where$\widehat{\ }$is the maximized log-likelihood.

We also consider the Cramér–von Mises (W^*) and Anderson–Darling (A^*) statistics. The statistics W^* and A^* are described in details in Chen and Balakrishnan (1995). In general, the smaller the values of these statistics, the better the fit to the data.

In Table 2, we list the numerical values of the statistics W^*, A^*, AIC, BIC, $HQIC$ and $CAIC$ using the two real data sets (DS), whereas the MLEs and their corresponding standard errors and the statistics of the model parameters are shown in Table 3. These numerical results are obtained using the `MATH-CAD` program.

Table 2. The statistics W^*, A^*, AIC, BIC, $HQIC$ and $CAIC$ for the two data sets

DS	Models	W^*	A^*	AIC	BIC	$HQIC$	$CAIC$
	TMOFr	0.2376	1.26771	309.973	320.393	314.19	310.394
	BFr	0.25137	1.39536	311.133	321.553	315.35	311.554
	GEFr	0.25872	1.43853	311.96	332.381	316.178	312.381
I	MOFr	0.59267	3.38252	351.328	359.143	354.491	351.578
	TFr	0.55598	3.17823	350.475	358.29	353.638	350.725
	Fr	0.54849	3.13643	348.308	353.519	350.417	348.432
	TMOFr	0.56541	3.10166	56.46	65.032	59.831	57.149
	MOFr	0.59607	3.2897	57.08	63.509	59.609	57.487
II	BFr	0.76879	4.20206	68.63	77.202	72.002	69.32
	GEFr	0.78121	4.27204	69.557	78.13	72.929	70.247
	TFr	1.17022	6.45074	100.078	106.507	102.606	100.484
	Fr	1.16252	6.40749	97.707	101.993	99.392	97.907

The figures in Table 2 indicate that the TMOFr model has the smallest values of the statistics W^*, A^*, AIC, BIC, $HQIC$ and $CAIC$ except BIC and $HQIC$ for the second data set. Hence, it can be chosen as the best model among all fitted models. Based on these criteria in Table 2, we conclude that the TMOFr distribution provides a better fit than the other models.

9. Concluding Remarks

In this paper, we propose a new four-parameter model, called the *transmuted Marshall-Olkin Fréchet* (TMOFr) distribution, which extends the Marshall-Olkin Fréchet (MOFr) distribution introduced by Krishna et al. (2013). In fact, the TMOFr distribution is motivated by the wide use of the Fréchet distribution in practice and also in view of the fact that the generalization provides more flexibility to analyze real life data. We provide some of its mathematical properties. The density function of TMOFr can be expressed as a mixture of Fréchet densities. We derive explicit expressions for the ordinary and incomplete moments, residual life and reversed residual life functions, quantile and generating functions, Rényi and q-entropies. We obtain the density function of order statistics and their moments. We discuss the maximum likelihood estimation of the model parameters. Two applications illustrate that the proposed distribution provides consistently better fit than other non-nested models.

Appendix A

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

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

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

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

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

Table 3. MLEs and their standard errors (in parentheses) for the two data sets

DS	Models	E			stimates
	TMOFr($\beta, \sigma, \alpha, \lambda$)	3.3313	0.6496	101.923	0.2936
		(0.206)	(0.068)	(47.625)	(0.27)
	BFr(β, σ, a, b)	0.4046	1.6097	22.0143	29.7617
		(0.108)	(2.498)	(21.432)	(17.479)
	GEFr(β, σ, a, b)	0.4776	1.3692	27.6452	17.4581
I		(0.133)	(2.017)	(14.136)	(14.818)
	MOFr(β, σ, α)	1.5796	2.3066	0.5988	
		(0.16)	(0.489)	(0.3091)	
	TFr(β, σ, λ)	1.7435	1.9315	0.0819	
		(0.076)	(0.097)	(0.198)	
	Fr(β, σ)	1.7766	1.8705		
		(0.113)	(0.112)		
	TMOFr($\beta, \sigma, \alpha, \lambda$)	6.8744	0.65	376.268	0.1499
		(0.596)	(0.049)	(246.832)	(0.302)
	MOFr(β, σ, α)	6.4655	0.6812	161.6114	
		(0.559)	(0.045)	(91.499)	
	BFr(β, σ, a, b)	0.6466	2.0518	15.0756	36.9397
II		(0.163)	(0.986)	(12.057)	(22.649)
	GEFr(β, σ, a, b)	0.7421	1.6625	32.112	13.2688
		(0.197)	(0.952)	(17.397)	(9.967)
	TFr(β, σ, λ)	2.7898	1.3068	0.1298	
		(0.165)	(0.034)	(0.208)	
	Fr(β, σ)	2.8876	1.2643		
		(0.234)	(0.059)		

References

Afify, A. Z., Nofal, Z. M., & Butt, N. S. (2014). Transmuted complementary Weibull geometric distribution. *Pakistan Journal of Statistics and Operation Research, 10*, 435-454. http://dx.doi.org/10.18187/pjsor.v10i4.836

Afify, A. Z., Yousof, H. M., Nofal, Z. M., & Cordeiro, G. M. (2015). The Weibull Fréchet distribution and its applications. Submitted.

Afify, A. Z., Nofal, Z. M., Yousof, H. M., El Gebaly, Y. M., & Butt, N. S. (2015). The transmuted Weibull Lomax distribution: properties and application. *Pakistan Journal of Statistics and Operation Research, 11*, 135-152. http://dx.doi.org/10.18187/pjsor.v11i1.956

Ahmad, A., Ahmad, S. P., & Ahmed, A. (2014). Transmuted inverse Rayleigh distribution: a generalization of the inverse Rayleigh distribution. *Mathematical Theory and Modeling, 4*, 90-98.

Ahmad, I., Kayid, M., & Pellerey, F. (2005). Further results involving the MIT order and the IMIT class. *Probability in the Engineering and Informational Science, 19,* 377-395. http://dx.doi.org/10.1017/s0269964805050229

Aryal, G. R., & Tsokos, C. P. (2009). On the transmuted extreme value distribution with application. *Nonlinear Analysis: Theory, Methods and Applications, 71,* 1401-1407. http://dx.doi.org/10.1016/j.na.2009.01.168

Aryal, G. R., & Tsokos C. P.(2011). Transmuted Weibull distribution: A Generalization of the Weibull Probability Distribution. *European Journal of Pure and Applied Mathematics, 4*, 89-102.

Barreto-Souza, W. M., Cordeiro, G. M., & Simas, A. B. (2011). Some results for beta Fréchet distribution. *Communications in Statistics-Theory and Methods, 40*, 798-811. http://dx.doi.org/10.1080/03610920903366149

Bourguignon, M., Ghosh, I., & Cordeiro, G. M. (2015). General results for the transmuted family of distributions and new models. Submitted.

Chen, G., Balakrishnan, N. (1995). A general purpose approximate goodness-of-fit test, *Journal of Quality Tech-*

nology, *27*, 154-161.

Elbatal, I. Asha, G., & Raja, V. (2014). transmuted exponentiated Fréchet distribution: properties and applications. *J. Stat. Appl. Prob.*, *3*, 379-394.

Fréchet, M. (1924). Sur la loi des erreurs d'observation, *Bulletin de la Société Mathématique de Moscou*, *33*, p. 5-8.

Glänzel, W. (1987). A characterization theorem based on truncated moments and its application to some distribution families. *Mathematical Statistics and Probability Theory (Bad Tatzmannsdorf, 1986), B, Reidel*, Dordrecht, 75-84. http://dx.doi.org/10.1007/978-94-009-3965-3_8

Glänzel, W. (1990). Some consequences of a characterization theorem based on truncated moments. *Statistics: A Journal of Theoretical and Applied Statistics*, *21*, 613-618. http://dx.doi.org/10.1080/02331889008802273

Guess, F., & Proschan, F. (1988). Mean residual life, theory and applications. In: Krishnaiah, P. R., Rao, C. R. (Eds.), *Handbook of Statistics. Reliability and Quality Control*, *7*, 215–224. http://dx.doi.org/10.1016/S0169-7161(88)07014-2

Gupta, R.C. (1975). On characterization of distributions by conditional expectations. *Communications in Statistics-Theory and Methods*, *4*, 99-103. http://dx.doi.org/10.1080/03610927508827230

Kayid, M., & Ahmad, I. (2004). On the mean inactivity time ordering with reliability applications, *Probability in the Engineering and Informational Science*, *18*, 395-409. http://dx.doi.org/10.1017/S0269964804183071

Keller, A. Z., & Kamath, A. R. (1982). Reliability analysis of CNC Machine Tools. *Reliability Engineering*, *3*, 449-473. http://dx.doi.org/10.1016/0143-8174(82)90036-1

Khan, M. S., & King, R. (2013). Transmuted Modified Weibull Distribution: A Generalization of the Modified Weibull Probability Distribution. *European Journal of Pure and Applied Mathematics*, *6*, 66-88.

Kilbas, A. A., Srivastava, H. M., & Trujillo, J. J. (2006). Theory and applications of fractional differential equations. Elsevier, Amsterdam.

Krishna, E., Jose, K. K., Alice, T., & Ristić, M. M. (2013). The Marshall-Olkin Fréchet Distribution. *Communications in Statistics-Theory and Methods*, *42*, 4091-4107. http://dx.doi.org/10.1080/03610926.2011.648785

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

Kotz, S., & Shanbhag, D. N. (1980). Some new approaches to probability distributions. *Adv. Appl. Prob.*, *12*, 903-921. http://dx.doi.org/10.2307/1426748

Lai, C. D., & Xie, M. (2006). *Stochastic aging and dependence for reliability*. Springer, New York.

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

Marshall, A. W., & 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. http://dx.doi.org/10.1093/biomet/84.3.641

Mead, M. E., & Abd-Eltawab A. R. (2014). A note on Kumaraswamy-Fréchet Distribution. *Aust. J. Basic and Appl. Sci.*, *8,* 294-300.

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

Nadarajah, S., & Kotz, S. (2003). The exponentiated exponential distribution, Available online at *http://interstat.statjournals.net/YEAR/2003/abstracts/0312001.php.*

Navarro, J. Franco, M., & Ruiz, J. M. (1998). Characterization through moments of the residual life and conditional spacing. *Sankhya:The Indian Journal of Statistics*, *60,* Series A, 36-48.

Nichols, M.D, Padgett, W.J. (2006). A Bootstrap control chart for Weibull percentiles. *Quality and Reliability Engineering International*, *22*, 141-151. http://dx.doi.org/10.1002/qre.691

Oguntunde, P. E., & Adejumo, A. O. (2015). The Transmuted Inverse Exponential Distribution. *International*

Journal of Advanced Statistics and Probability, *3*, 1-7. http://dx.doi.org/10.14419/ijasp.v3i1.3684

Shaw, W. T., & Buckley, I. R. C. (2007). *The alchemy of probability distributions: beyond Gram-Charlier expansions and a skew-kurtotic-normal distribution from a rank transmutation map*. Research report.

Silva, R. V. D., de Andrade, T. A., Maciel, D., Campos, R. P., and Cordeiro, G. M. (2013). A new lifetime model: The gamma extended Fréchet distribution. *Journal of Statistical Theory and Applications*, *12*, 39-54. http://dx.doi.org/10.2991/jsta.2013.12.1.4

Smith, R. L., & Naylor, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Applied Statistics*, *36*, 358-369. http://dx.doi.org/10.2307/2347795

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

Zoroa, P., Ruiz, J. M., & Marin, J. (1990). A characterization based on conditional expectations. *Communications in Statistics-Theory Methods*, *19*, 3127-3135. http://dx.doi.org/10.1080/03610929008830368

International Journal of Statistics and Probability; Vol. 4, No. 4; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Reviewer Acknowledgements

International Journal of Statistics and Probability wishes to acknowledge the following individuals for their assistance with peer review of manuscripts for this issue. Their help and contributions in maintaining the quality of the journal is greatly appreciated.

Many authors, regardless of whether *International Journal of Statistics and Probability* publishes their work, appreciate the helpful feedback provided by the reviewers.

Reviewers for Volume 4, Number 4

Abdullah A. SMADI

Afsin Sahin

Ali Reza Fotouhi

Anwar H Joarder

Bibi Abdelouahab

Carolyn Huston

Douglas Lorenz

Encarnación Alvarez-Verdejo

Gabriel A. Okyere

Gane Samb Lo

Hongsheng Dai

Ivair R. Silva

Marcelo Bourguignon

Milind Phadnis

Mirko D'Ovidio

Philip Westgate

Rebecca Bendayan

Sajid Ali

Samir Safi

Sohair F. Higazi

Tewfik Kernane

Vyacheslav Abramov

Wei Zhang

Wojciech Gamrot

Yimei Li

Wendy Smith

On behalf of,

The Editorial Board of *International Journal of Statistics and Probability*

Canadian Center of Science and Education

Call for Manuscripts

International Journal of Statistics and Probability is a peer-reviewed journal, published by Canadian Center of Science and Education. The journal publishes research papers in all areas of statistics and probability. The journal is available in electronic form in conjunction with its print edition. All articles and issues are available for free download online.

We are seeking submissions for forthcoming issues. All manuscripts should be written in English. Manuscripts from 3000–8000 words in length are preferred. All manuscripts should be prepared in LaTeX or MS-Word format, and submitted online, or sent to: ijsp@ccsenet.org

Paper Selection and Publishing Process

a) Upon receipt of a submission, the editor sends an e-mail of confirmation to the submission's author within one to three working days. If you fail to receive this confirmation, your submission e-mail may have been missed.

b) Peer review. We use a double-blind system for peer review; both reviewers' and authors' identities remain anonymous. The paper will be reviewed by at least two experts: one editorial staff member and at least one external reviewer. The review process may take two to three weeks.

c) Notification of the result of review by e-mail.

d) If the submission is accepted, the authors revise paper and pay the publication fee.

e) After publication, the corresponding author will receive two hard copies of the journal, free of charge. If you want to keep more copies, please contact the editor before making an order.

f) A PDF version of the journal is available for download on the journal's website, free of charge.

Requirements and Copyrights

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the authorities responsible where the work was carried out, and that, if accepted, the article will not be published elsewhere in the same form, in English or in any other language, without the written consent of the publisher. The editors reserve the right to edit or otherwise alter all contributions, but authors will receive proofs for approval before publication.

Copyrights for articles are retained by the authors, with first publication rights granted to the journal. The journal/publisher is not responsible for subsequent uses of the work. It is the author's responsibility to bring an infringement action if so desired by the author.

More Information

E-mail: ijsp@ccsenet.org

Website: www.ccsenet.org/ijsp

Paper Submission Guide: www.ccsenet.org/submission

Recruitment for Reviewers: www.ccsenet.org/reviewer

The journal is peer-reviewed
The journal is open-access to the full text
The journal is included in:

BASE
EBSCOhost
Google Scholar
JournalTOCs
Library and Archives Canada
LOCKSS
PKP Open Archives Harvester
ProQuest
SHERPA/RoMEO
Standard Periodical Directory
Ulrich's

International Journal of Statistics and Probability

Quarterly

Publisher	Canadian Center of Science and Education
Address	1120 Finch Avenue West, Suite 701-309, Toronto, ON., M3J 3H7, Canada
Telephone	1-416-642-2606
Fax	1-416-642-2608
E-mail	ijsp@ccsenet.org
Website	www.ccsenet.org/ijsp

ISSN 1927-7032