ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Volume 16, No.4
October 2018

REVSTAT

Statistical Journal

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística = ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: + 351 21 842 61 00
    Fax: + 351 21 845 40 84
    Web site: http://www.ine.pt
    Customer Support Service
    + 351 218 440 695
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *140 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*
- **PRICE [VAT included]**
  - *€ 9,00*

INDEX

Abstracted/indexed in: *Current Index to Statistics, DOAJ, Google Scholar, Journal Citation Reports/Science Edition, Mathematical Reviews, Science Citation Index Expanded®, SCOPUS* and *Zentralblatt für Mathematic.*

REVSTAT – Statistical Journal
Volume 16, Number 4, October 2018, 405–428

COMPARISON OF THE AVERAGE KAPPA COEFFICIENTS OF BINARY DIAGNOSTIC TESTS DONE ON THE SAME SUBJECTS

Authors: José Antonio Roldán-Nofuentes
– Statistics (Biostatistics), School of Medicine, University of Granada, Granada, Spain
jaroldan@ugr.es

Carmen Olvera-Porcel
– Statistics (Biostatistics), School of Medicine, University of Granada, Granada, Spain mcolvera@ugr.es

Received: October 2015 Revised: September 2016 Accepted: September 2016

Abstract:

- The average kappa coefficient of a binary diagnostic test is a chance corrected index between the binary diagnostic test and the gold standard, and it depends on the sensitivity and the specificity of the diagnostic test and on the disease prevalence. In this article, several hypothesis tests are studied to compare the average kappa coefficients of two (o more) binary diagnostic tests done on the same subjects. Simulation experiments were carried out to study the type I errors and the powers of the hypothesis tests studied. A program in R was written to solve the problem studied and it can be freely downloaded from the Internet. The results were applied to a real example on the diagnosis of coronary disease.

Key-Words:

- *average kappa coefficient; binary diagnostic test; weighted kappa coefficient.*

AMS Subject Classification:

- 62P10, 62H15.

1. INTRODUCTION

The fundamental parameters to assess and compare the performance of binary diagnostic tests are sensitivity and specificity. Sensitivity is the probability of the result of the binary diagnostic test (BDT) being positive when the individual has the disease, and specificity is the probability of the result of the BDT being negative when the individual does not have the disease. Both parameters depend only on the specific characteristics of the BDT, i.e. the intrinsic properties (physical, biological, chemical, etc.) of the BDT. When comparing two BDTs in paired designs, i.e. when the two BDTs and the gold standard (GS) are applied to all of the individuals in a random sample, the comparison of the two sensitivities (specificities) is made conditioning in the total of individuals with the disease (without the disease) and applying the exact test to compare two binomial proportions or its asymptotic version (McNemar's test).

When considering the losses associated with an erroneous classification with the BDT, the parameter that is used to assess the BDT is the weighted kappa coefficient [1,2,3]. The weighted kappa coefficient depends on the sensitivity and the specificity of the BDT, on the disease prevalence in the population studied and on the relative loss between the false positives and the false negatives (weighting index). The value of the weighting index is set by the clinical laboratory researcher based on his or her knowledge about the problem to be solved. Bloch [4] studied the comparison of the weighted kappa coefficients of two BDTs in relation to the same GS subject to a paired design.

The problem posed by the weighted kappa coefficient as a measure to assess and compare the performance of BDTs is the allocation of the value to the weighting index, since the clinical laboratory researcher does not have enough knowledge about the problem to be able to allocate that value, and two clinicians might even allocate different values to that index in the same problem. In order to solve this problem, Roldán-Nofuentes and Olvera-Porcel [5] defined a new parameter called the average kappa coefficient. The average kappa coefficient of the BDT depends on the sensitivity and the specificity of the BDT and on the disease prevalence, and does not depend on the weighting index. This new parameter has properties that make it valid to assess and compare BDTs. In this study, several hypothesis tests are studied to compare the average kappa coefficients of two BDTs in a paired design. In Section 2, the weighted kappa coefficient and the average kappa coefficient are explained. In Section 3, we present several asymptotic hypothesis tests to compare the average kappa coefficients of two BDTs subject to paired design. In Section 4, simulation experiments are carried out to study the type I errors and the powers of the hypothesis tests presented in Section 3. In Section 5, we study the situation in which more than two BDTs are compared. In Section 6, we present a program written in R which allows us to solve the problem posed. In Section 7, the results obtained are applied to a real example, and in Section 8 the results obtained are discussed.

2. WEIGHTED KAPPA COEFFICIENT AND AVERAGE KAPPA COEFFICIENT

Let L and L' be the losses associated with an erroneous classification with the BDT: L is the loss that occurs when for an individual the BDT is negative and the GS is positive, and L' is the loss that occurs when the BDT is positive and the GS is negative. Losses L and L' are zero when an individual (with or without the disease) is classified correctly with the BDT. The weighted kappa coefficient of a BDT is [1,2,3,4,6]

$$\kappa(c) = \frac{pq(Se + Sp - 1)}{p(1-Q)c + qQ(1-c)},$$

where Se is the sensitivity of the BDT, Sp the specificity, p the disease prevalence, $q = 1-p$, $Q = pSe + q(1-Sp)$ and $c = L/(L+L')$ is the weighting index. When loss L is equal to zero then $c = 0$, and the weighted kappa coefficient is

$$\kappa(0) = \frac{Sp - (1-Q)}{Q},$$

and when loss L' is equal to zero then $c = 1$, and the weighted kappa coefficient is

$$\kappa(1) = \frac{Se - Q}{1-Q}.$$

The weighted kappa coefficient can also be written as

$$\kappa(c) = \frac{p(1-Q)c\kappa(1) + qQ(1-c)\kappa(0)}{p(1-Q)c + qQ(1-c)},$$

and therefore it is a weighted mean of $\kappa(0)$ and $\kappa(1)$. Weighting index c varies between 0 and 1 and represents the relative loss between the false positives and the false negatives. In practice the weighting index c is unknown, but its values can be assumed according to the objective for which the diagnostic test is going to be used. If the diagnostic test is going to be used as a previous step for a risky treatment (e.g. surgery), there is more concern about the false positives and the c index is lower than 0.5; if the diagnostic test is going to be used as a screening test, there is more concern about the false negatives and the c index is greater than 0.5; and the c index is 0.5 when the diagnostic test is used for a simple diagnosis. If $L = L'$, then $c = 0.5$ and $\kappa(0.5)$ is called the Cohen kappa coefficient; if $L > L'$, then $0.5 < c < 1$, and if $L' > L$ then $0 < c < 0.5$. The properties of the weighted kappa coefficient can be seen in the manuscript of Kraemer *et al.* [3] and in that of Roldán-Nofuentes *et al.* [6]. The problem posed by the weighted kappa coefficient is the allocation of a value to the weighting index. Allocating values 0 or 1 means that one of the losses is equal to zero, which is not realistic. In practice, the allocation is made based on the knowledge that the clinical laboratory researcher has about the problem that is being analyzed. This procedure can lead to some disagreement, since two different clinicians may allocate different values and their conclusions may not be the same.

In order to solve this problem of the allocation of values to the weighting index, Roldán-Nofuentes and Olvera-Porcel [5] defined a new parameter: the average kappa coefficient. The average kappa coefficient is a measure of the weighted kappa coefficients, and only depends on the sensitivity and the specificity of the BDT and the disease prevalence, and does not depend on the weighting index. If the clinical laboratory researcher considers that the loss associated with a false positive is greater than the loss associated with a false negative, $L' > L$ and $0 < c < 0.5$, the average kappa coefficient is

$$\kappa_1 = \frac{1}{0.5}\int_0^{0.5} \kappa(c)\, dc = \begin{cases} \frac{2\kappa(0)\kappa(1)}{\kappa(0)-\kappa(1)} \ln\left[\frac{\kappa(0)+\kappa(1)}{2\kappa(1)}\right], & p \neq Q \\ Se + Sp - 1, & p = Q, \end{cases} \tag{2.1}$$

i.e. the average kappa coefficient κ_1 is the average value of $\kappa(c)$ when $0 < c < 0.5$. If the clinical laboratory researcher considers that the loss associated with a false negative is greater than the loss associated with a false positive, $L > L'$ and $0.5 < c < 1$, the average kappa coefficient is

$$\kappa_2 = \frac{1}{0.5}\int_{0.5}^{1} \kappa(c)\, dc = \begin{cases} \frac{2\kappa(0)\kappa(1)}{\kappa(0)-\kappa(1)} \ln\left[\frac{2\kappa(0)}{\kappa(0)+\kappa(1)}\right], & p \neq Q \\ Se + Sp - 1, & p = Q. \end{cases} \tag{2.2}$$

As the weighted kappa coefficient is a measure of the beyond-chance agreement between the BDT and the GS, then κ_1 and κ_2 are measures of the average beyond-chance agreement between the BDT and the GS. The properties of κ_1 and κ_2 can be seen in the manuscript by Roldán-Nofuentes and Olvera-Porcel [5], and they are parameters that allow us to assess and compare the performance of BDTs. The comparison of the average kappa coefficients of two BDTs subject to paired design is now studied.

3. COMPARISON OF TWO AVERAGE KAPPA COEFFICIENTS

Let us consider two BDTs that are compared in relation to the same GS. The frequencies obtained applying the two BDTs and the GS to a sample of n individuals and theoretical probabilities are shown in Table 1, where the variable T_i models the result of the i-th BDT ($T_i = 1$ when the result is positive and $T_i = 0$ when it is negative) and the variable D models the result of the GS ($D = 1$ when the individual has the disease and $D = 0$ when this is not the case). If the clinical laboratory researcher assumes a value for the weighting index c, Bloch [4] has studied the comparison of the weighted kappa coefficients of two BDTs subject to a paired design. Using the notation in Table 1, the estimators of the weighted kappa coefficients deduced by Bloch [4] are

$$\hat{\kappa}_1(c) = \frac{(s_{11}+s_{10})(r_{01}+r_{00}) - (s_{01}+s_{00})(r_{10}+r_{11})}{sc\sum_{k=0}^{1}(s_{0k}+r_{0k}) + r(1-c)\sum_{k=0}^{1}(s_{1k}+r_{1k})}$$

and

$$\hat{\kappa}_2(c) = \frac{(s_{11}+s_{01})(r_{10}+r_{00}) - (s_{10}+s_{00})(r_{01}+r_{11})}{sc\sum_{h=0}^{1}(s_{h0}+r_{h0}) + r(1-c)\sum_{h=0}^{1}(s_{h1}+r_{h1})},$$

and the statistic for $H_0 : \kappa_1(c) = \kappa_2(c)$ vs $H_1 : \kappa_1(c) \neq \kappa_2(c)$ is

$$z = \frac{\hat{\kappa}_1(c) - \hat{\kappa}_2(c)}{\sqrt{\hat{V}ar[\hat{\kappa}_1(c)] + \hat{V}ar[\hat{\kappa}_2(c)] - 2\hat{C}ov[\hat{\kappa}_1(c), \hat{\kappa}_2(c)]}} \xrightarrow[n\to\infty]{} N(0,1),$$

where the expressions of the variances and the covariance have been obtained by Bloch [4] applying the delta method.

Table 1: Observed frequencies and probabilities subject to paired design.

Observed frequencies					
	$T_1=1$		$T_1=0$		Total
	$T_2=1$	$T_2=0$	$T_2=1$	$T_2=0$	
$D=1$	s_{11}	s_{10}	s_{01}	s_{00}	s
$D=0$	r_{11}	r_{10}	r_{01}	r_{00}	r
Total	n_{11}	n_{10}	n_{01}	n_{00}	n

Probabilities					
	$T_1=1$		$T_1=0$		Total
	$T_2=1$	$T_2=0$	$T_2=1$	$T_2=0$	
$D=1$	p_{11}	p_{10}	p_{01}	p_{00}	p
$D=0$	q_{11}	q_{10}	q_{01}	q_{00}	q
Total	$p_{11}+q_{11}$	$p_{10}+q_{10}$	$p_{01}+q_{01}$	$p_{00}+q_{00}$	1

We then study the comparison of the average kappa coefficients of the two BDTs. Firstly, we study the comparison of the two average kappa coefficients when the clinical laboratory researcher considers that $L' > L$ $(0 < c < 0.5)$ and after when $L > L'$ $(0.5 < c < 1)$.

When $L' > L$ the hypothesis test to compare the two average kappa coefficients is $H_0 : \kappa_{11} = \kappa_{21}$ vs $H_1 : \kappa_{11} \neq \kappa_{21}$, where κ_{i1} is the average kappa coefficient of the i-th BDT when the clinical laboratory researcher considers that $L' > L$. In terms of the probabilities in Table 1, the sensitivity and the specificity of each BDT are written as $Se_1 = (p_{10}+p_{11})/p$, $Sp_1 = (q_{00}+q_{01})/q$, $Se_2 = (p_{01}+p_{11})/p$ and $Sp_2 = (q_{00}+q_{10})/q$, where $p = \sum_{ij} p_{ij}$ is the disease prevalence and $q = 1-p = \sum_{ij} q_{ij}$. Replacing in equation (2.1) each parameter with its ex-

pression, the average kappa coefficient κ_{11} is written as

$$\begin{aligned}\kappa_{11} &= \frac{2\kappa_1(0)\,\kappa_1(1)}{\kappa_1(0)-\kappa_1(1)}\,\ln\left\{\frac{\kappa_1(0)+\kappa_1(1)}{2\kappa_1(1)}\right\}\\ &= 2\left(\frac{\sum_{j=0}^{1}(p_{0j}+q_{0j})}{\frac{1}{p}\sum_{j=0}^{1}p_{1j}-\sum_{j=0}^{1}(p_{1j}+q_{1j})}-\frac{\sum_{j=0}^{1}(p_{1j}+q_{1j})}{\frac{1}{q}\sum_{j=0}^{1}q_{0j}-\sum_{j=0}^{1}(p_{0j}+q_{0j})}\right)^{-1}\\ &\quad\times\ln\left[\frac{1}{2}\left(\frac{\left(\sum_{j=0}^{1}(p_{0j}+q_{0j})\right)\left(\frac{1}{q}\sum_{j=0}^{1}q_{0j}-\sum_{j=0}^{1}(p_{0j}+q_{0j})\right)}{\left(\sum_{j=0}^{1}(p_{1j}+q_{1j})\right)\left(\frac{1}{p}\sum_{j=0}^{1}p_{1j}-\sum_{j=0}^{1}(p_{1j}+q_{1j})\right)}+1\right)\right]\end{aligned}$$

when $p \neq Q_1$ and $\kappa_{11} = \frac{1}{p}\sum_{j=0}^{1}p_{1j} + \frac{1}{q}\sum_{j=0}^{1}q_{0j} - 1$ when $p = Q_1$. Regarding κ_{21}, its expression is

$$\begin{aligned}\kappa_{21} &= \frac{2\kappa_2(0)\,\kappa_2(1)}{\kappa_2(0)-\kappa_2(1)}\,\ln\left\{\frac{\kappa_2(0)+\kappa_2(1)}{2\kappa_2(1)}\right\}\\ &= 2\left(\frac{\sum_{i=0}^{1}(p_{i0}+q_{i0})}{\frac{1}{p}\sum_{i=0}^{1}p_{i1}-\sum_{i=0}^{1}(p_{i1}+q_{i1})}-\frac{\sum_{i=0}^{1}(p_{i1}+q_{i1})}{\frac{1}{q}\sum_{i=0}^{1}q_{i0}-\sum_{i=0}^{1}(p_{i0}+q_{i0})}\right)^{-1}\\ &\quad\times\ln\left[\frac{1}{2}\left(\frac{\left(\sum_{i=0}^{1}(p_{i0}+q_{i0})\right)\left(\frac{1}{q}\sum_{i=0}^{1}q_{i0}-\sum_{i=0}^{1}(p_{i0}+q_{i0})\right)}{\left(\sum_{i=0}^{1}(p_{i1}+q_{i1})\right)\left(\frac{1}{p}\sum_{i=0}^{1}p_{i1}-\sum_{i=0}^{1}(p_{i1}+q_{i1})\right)}+1\right)\right]\end{aligned}$$

when $p \neq Q_2$ and $\kappa_{21} = \frac{1}{p}\sum_{j=0}^{1}p_{1j} + \frac{1}{q}\sum_{j=0}^{1}q_{0j} - 1$ when $p = Q_2$. As the probabilities p_{ij} and q_{ij} are probabilities of a multinomial distribution, their estimators are $\hat{p}_{ij} = s_{ij}/n$ and $\hat{q}_{ij} = r_{ij}/n$. Therefore, the estimator of κ_{11} is

$$\begin{aligned}\hat{\kappa}_{11} &= \frac{2\left\{(s_{10}+s_{11})(r_{00}+r_{01})-(s_{00}+s_{01})(r_{10}+r_{11})\right\}}{n\left(\sum_{j=0}^{1}(s_{0j}-r_{1j})\right)}\\ &\quad\times\ln\left[\frac{1}{2}\left(\frac{s\sum_{j=0}^{1}(s_{0j}+r_{0j})}{r\sum_{j=0}^{1}(s_{1j}+r_{1j})}+1\right)\right]\end{aligned}$$

when $\hat{p} \neq \hat{Q}_1$, i.e. when $s_{01}+s_{00} \neq r_{10}+r_{11}$, and

$$\hat{\kappa}_{11} = \frac{(s_{10}+s_{11})(r_{00}+r_{01})-(s_{00}+s_{01})(r_{10}+r_{11})}{sr}$$

when $\hat{p} = \hat{Q}_1$, i.e. when $s_{01} + s_{00} = r_{10} + r_{11}$. Regarding the estimator of κ_{21}, its expression is

$$\hat{\kappa}_{21} = \frac{2\left\{(s_{01}+s_{11})(r_{00}+r_{10}) - (s_{00}+s_{10})(r_{01}+r_{11})\right\}}{n\left(\sum_{i=0}^{1}(s_{i0}-r_{i1})\right)} \times \ln\left[\frac{1}{2}\left(\frac{s\sum_{i=0}^{1}(s_{i0}+r_{i0})}{r\sum_{i=0}^{1}(s_{i1}+r_{i1})}+1\right)\right]$$

when $\hat{p} \neq \hat{Q}_2$, i.e. $s_{10} + s_{00} \neq r_{01} + r_{11}$, and

$$\hat{\kappa}_{21} = \frac{(s_{01}+s_{11})(r_{00}+r_{10}) - (s_{00}+s_{10})(r_{01}+r_{11})}{sr}$$

when $\hat{p} = \hat{Q}_2$, i.e. $s_{10} + s_{00} = r_{01} + r_{11}$. Applying the delta method, the asymptotic variance-covariance matrix of $\hat{\kappa}_{11}$ and $\hat{\kappa}_{21}$ is

$$\sum\nolimits_{\hat{\boldsymbol{\kappa}}_1} = \left(\frac{\partial \boldsymbol{\kappa}_1}{\partial \boldsymbol{\pi}}\right) \sum\nolimits_{\hat{\boldsymbol{\pi}}} \left(\frac{\partial \boldsymbol{\kappa}_1}{\partial \boldsymbol{\pi}}\right)^T,$$

where $\boldsymbol{\kappa}_1 = (\kappa_{11}, \kappa_{21})^T$, $\boldsymbol{\pi} = (p_{11}, p_{10}, p_{01}, p_{00}, q_{11}, q_{10}, q_{01}, q_{00})^T$ and

$$\sum\nolimits_{\hat{\boldsymbol{\pi}}} = \frac{\mathrm{Diag}\,(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T}{n}$$

is the variance-covariance matrix of the probabilities in Table 1. Replacing in the expression of $\sum_{\hat{\boldsymbol{\kappa}}_1}$ each parameter with its estimator, we obtain the expressions of the estimated asymptotic variances-covariances of $\hat{\boldsymbol{\kappa}}_1$. These expressions are not presented here as they are very long and complicated (they were calculated using the R programming approach created to solve this hypothesis test). Finally, the statistic to contrast the equality of the average kappa coefficients when $L' > L$ is

$$z = \frac{\hat{\kappa}_{11} - \hat{\kappa}_{21}}{\sqrt{\hat{V}ar\,(\hat{\kappa}_{11}) + \hat{V}ar\,(\hat{\kappa}_{21}) - 2\hat{C}ov\,(\hat{\kappa}_{11}, \hat{\kappa}_{21})}} \xrightarrow[n\to\infty]{} N(0,1).$$

Furthermore, an asymptotic confidence interval for the difference of the average kappa coefficients is

$$\kappa_{11} - \kappa_{21} \in \hat{\kappa}_{11} - \hat{\kappa}_{21} \pm z_{1-\alpha/2}\sqrt{\hat{V}ar\,(\hat{\kappa}_{11}) + \hat{V}ar\,(\hat{\kappa}_{21}) - 2\hat{C}ov\,(\hat{\kappa}_{11}, \hat{\kappa}_{21})},$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution.

If the clinical laboratory researcher considers that $L > L'$, and therefore that $0.5 < c < 1$, the hypothesis test to compare the two average kappa coefficients is $H_0 : \kappa_{12} = \kappa_{22}$ vs $H_1 : \kappa_{12} \neq \kappa_{22}$, where κ_{i2} is the average kappa coefficient of

the i-th BDT when $L > L'$. The process to solve this hypothesis test is similar to the previous case, and the statistic is

$$z = \frac{\hat{\kappa}_{12} - \hat{\kappa}_{22}}{\sqrt{\hat{V}ar\left(\hat{\kappa}_{12}\right) + \hat{V}ar\left(\hat{\kappa}_{22}\right) - 2\hat{C}ov\left(\hat{\kappa}_{12}, \hat{\kappa}_{22}\right)}} \xrightarrow[n\to\infty]{} N\left(0,1\right).$$

Replacing in equation (2.2) each parameter with its expression, the estimators of κ_{12} is

$$\begin{aligned}\hat{\kappa}_{12} &= \frac{2\left\{\left(s_{10}+s_{11}\right)\left(r_{00}+r_{01}\right)-\left(s_{00}+s_{01}\right)\left(r_{10}+r_{11}\right)\right\}}{n\left(\sum_{j=0}^{1}\left(s_{0j}-r_{1j}\right)\right)} \\ &\times \ln\left[2\frac{s\sum_{j=0}^{1}\left(s_{0j}+r_{0j}\right)}{s\sum_{j=0}^{1}\left(s_{0j}+r_{0j}\right)+r\sum_{j=0}^{1}\left(s_{1j}+r_{1j}\right)}\right]\end{aligned}$$

when $\hat{p} \neq \hat{Q}_1$, i.e. $s_{01}+s_{00} \neq r_{10}+r_{11}$, and

$$\hat{\kappa}_{12} = \frac{\left(s_{10}+s_{11}\right)\left(r_{00}+r_{01}\right)-\left(s_{00}+s_{01}\right)\left(r_{10}+r_{11}\right)}{sr}$$

when $\hat{p} = \hat{Q}_1$, i.e. $s_{01}+s_{00} = r_{10}+r_{11}$. Regarding κ_{22}, it holds that

$$\begin{aligned}\hat{\kappa}_{22} &= \frac{2\left\{\left(s_{01}+s_{11}\right)\left(r_{00}+r_{10}\right)-\left(s_{00}+s_{10}\right)\left(r_{01}+r_{11}\right)\right\}}{n\left(\sum_{i=0}^{1}\left(s_{i0}-r_{i1}\right)\right)} \\ &\times \ln\left[2\frac{s\sum_{i=0}^{1}\left(s_{i0}+r_{i0}\right)}{s\sum_{i=0}^{1}\left(s_{i0}+r_{i0}\right)+r\sum_{i=0}^{1}\left(s_{i1}+r_{i1}\right)}\right]\end{aligned}$$

when $\hat{p} \neq \hat{Q}_2$, i.e. $s_{10}+s_{00} \neq r_{01}+r_{11}$, and

$$\hat{\kappa}_{22} = \frac{\left(s_{01}+s_{11}\right)\left(r_{00}+r_{10}\right)-\left(s_{00}+s_{10}\right)\left(r_{01}+r_{11}\right)}{sr}$$

when $\hat{p} = \hat{Q}_2$, i.e. $s_{10}+s_{00} = r_{01}+r_{11}$. The asymptotic variance-covariance matrix is estimated in a similar way to the previous case. Moreover, an asymptotic confidence interval for the difference of the average kappa coefficients is

$$\kappa_{12}-\kappa_{22} \in \hat{\kappa}_{12}-\hat{\kappa}_{22} \pm z_{1-\alpha/2}\sqrt{\hat{V}ar\left(\hat{\kappa}_{12}\right)+\hat{V}ar\left(\hat{\kappa}_{22}\right)-2\hat{C}ov\left(\hat{\kappa}_{12},\hat{\kappa}_{22}\right)}.$$

The comparison of the average kappa coefficients can also be made using transformations, such as the logarithm and the logit transformations. In this case, the hypothesis test is $H_0 : F\left(\kappa_{1k}\right) = F\left(\kappa_{2k}\right)$ vs $H_1 : F\left(\kappa_{1k}\right) \neq F\left(\kappa_{2k}\right)$, where F is the logarithm or the logit respectively. The problem is solved in a similar way to in the previous case. These transformations aim to improve the convergence of the distribution of the estimators to the normal distribution.

4. SIMULATION EXPERIMENTS

Simulation experiments were carried out to study the type I errors and the powers of the hypothesis tests $H_0 : \kappa_{1k} = \kappa_{2k}$ and $H_0 : F(\kappa_{1k}) = F(\kappa_{2k})$. Therefore, 5000 random samples of multinomial distributions were generated with sizes of 100, 200, 300, 400, 500, 1000 and 2000, which are sizes in a wide range to show the behaviour of the hypothesis tests. The probabilities of the multinomial distributions were calculated using the conditional dependence model proposed by Vacek [7], i.e.

$$p_{ij} = P(T_1 = i, T_2 = j \,|D = 1) = P(T_1 = i \,|D = 1) \times P(T_2 = j \,|D = 1) + \delta_{ij}\varepsilon_1$$

and

$$q_{ij} = P(T_1 = i, T_2 = j \,|D = 0) = P(T_1 = i \,|D = 0) \times P(T_2 = j \,|D = 0) + \delta_{ij}\varepsilon_0,$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = -1$ if $i \neq j$, and ε_i is the covariance between the two BDTs when $D = i$. Vacek [7] demonstrated that

$$0 \leq \varepsilon_1 \leq Min\{Se_1(1 - Se_2), (1 - Se_1)Se_2\}$$

and that

$$0 \leq \varepsilon_0 \leq Min\{Sp_1(1 - Sp_2), (1 - Sp_1)Sp_2\}.$$

If $\varepsilon_1 = \varepsilon_0 = 0$ then the two BDTs are conditionally independent on the disease status. In practice, the assumption of conditional independence is not very realistic and therefore $\varepsilon_1 > 0$ and/or $\varepsilon_0 > 0$.

The simulation experiments were designed based on the equations of the average kappa coefficients of the two BDTs, i.e.

$$\kappa_{i1} = \frac{2\kappa_i(0)\,\kappa_i(1)}{\kappa_i(0) - \kappa_i(1)} \ln\left[\frac{\kappa_i(0) + \kappa_i(1)}{2\kappa_i(1)}\right] \tag{4.1}$$

and

$$\kappa_{i2} = \frac{2\kappa_i(0)\,\kappa_i(1)}{\kappa_i(0) - \kappa_i(1)} \ln\left[\frac{2\kappa_i(0)}{\kappa_i(0) + \kappa_i(1)}\right]. \tag{4.2}$$

As the disease prevalence, we took the values 5%, 10%, 30% and 50%. The first two values correspond to a scenario with low prevalence and the last two with a high disease prevalence, and they are a range of values that allow us to study the effect of the prevalence on the behaviour of each hypothesis test. Regarding the average kappa coefficients we took the values 0.2, 0.4, 0.6 and 0.8. Therefore, following the idea of Cicchetti [8] we took values of average kappa coefficients with different levels of significance: poor (< 0.40), fair ($0.40 - 0.59$), good ($0.60 - 0.74$) and excellent ($0.75 - 1$). Once the values for the prevalence and the average kappa coefficient were set, using the Newton–Raphson method, the system made up of equations (4.1) and (4.2) was solved to thus obtain the values of $\kappa_i(0)$ and $\kappa_i(1)$,

only considering those values whose solutions are between 0 and 1. Finally, in order to obtain the values of the sensitivity and the specificity of each BDT (Se_i and Sp_i) the system made up of the equations $\kappa_i(0) = \{Sp_i - (1 - Q_i)\}/Q_i$ and $\kappa_i(1) = (Se_i - Q_i)/(1 - Q_i)$ was solved. Once the values for Se_i and Sp_i were obtained, the maximum values for the covariances ε_1 and ε_0 were calculated. Finally, the probabilities of the multinomial distributions were calculated based on the model proposed by Vacek [7]. Furthermore, the samples were generated in such a way that in all cases it was possible to estimate all of the parameters and their variances-covariances. In all of the study, we took as the nominal error $\alpha = 5\%$. In the tables with results, Test 1 refers to the hypothesis test without logarithmic transformation and Test 2 refers to the hypothesis test with logarithmic transformation. The results with the logit transformation are not shown as they are very similar to those obtained with the logarithmic transformation.

4.1. Type I errors

In Table 2, we can see some of the results obtained for the type I errors of the hypothesis tests $H_0 : \kappa_{11} = \kappa_{21}$ (Test 1) and $H_0 : \ln(\kappa_{11}) = \ln(\kappa_{21})$ (Test 2), i.e. when comparing the average kappa coefficients considering that $L' > L$. In Table 3, we can see some results for the type I errors of the hypothesis test $H_0 : \kappa_{12} = \kappa_{22}$ (Test 1) and $H_0 : \ln(\kappa_{12}) = \ln(\kappa_{22})$ (Test 2), i.e. when comparing the average kappa coefficients considering that $L > L'$. In these tables we can see the values of the sensitivities, specificities, prevalence and covariances with which the multinomial samples were generated.

When $L' > L$ (Table 2), the disease prevalence and the covariances between the two BDTs have an important effect upon the type I error of the test $H_0 : \kappa_{11} = \kappa_{21}$. The increase in the prevalence implies an increase in the type I error, especially in samples of 100 and 200, although without overwhelming the nominal error (a situation which has been considered when the type I error is greater than 6.5%). The increase in the values of the covariances implies a decrease in the type I error, especially for $n \leq 500$. In general terms, when the values of the covariances are high, the hypothesis test $H_0 : \kappa_{11} = \kappa_{21}$ is conservative (its type I error is lower than the nominal error) for a sample size $n \leq 500$ (depending on the disease prevalence). The prevalence and the covariances have practically no effect upon the type I error when the samples are very large ($n \geq 1000$). Therefore, in general terms, the type I error of the test $H_0 : \kappa_{11} = \kappa_{21}$ is lower than the nominal error and starting from a certain sample size it fluctuates around the nominal error without overwhelming it. Regarding the type I error of the test $H_0 : \ln(\kappa_{11}) = \ln(\kappa_{21})$, its behavior is, in general terms, very similar to that of the test $H_0 : \kappa_{11} = \kappa_{21}$, although for sample sizes of 100 and 200 its type I error is somewhat lower than that of the hypothesis test without transformation.

Table 2: Type I errors of the hypothesis tests when $L' > L$.

$\kappa_{11} = \kappa_{21} = 0.2$ $Se_1 = 0.7773,\ Sp_1 = 0.7308,\ Se_2 = 0.7773,\ Sp_2 = 0.7308$ $p = 10\%,\ \varepsilon_1 \leq 0.1731,\ \varepsilon_0 \leq 0.1967$						
$\boldsymbol{n}$	$\varepsilon_1 = 0, \varepsilon_0 = 0$		$\varepsilon_1 = 0.08, \varepsilon_0 = 0.09$		$\varepsilon_1 = 0.16, \varepsilon_0 = 0.18$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.022	0.009	0.012	0.008	0	0
200	0.044	0.026	0.031	0.022	0.001	0
300	0.047	0.040	0.035	0.029	0.004	0.004
400	0.045	0.040	0.050	0.042	0.004	0.004
500	0.050	0.048	0.044	0.042	0.010	0.008
1000	0.048	0.046	0.047	0.046	0.020	0.020
2000	0.055	0.056	0.056	0.055	0.044	0.043

$\kappa_{11} = \kappa_{21} = 0.4$ $Se_1 = 0.8864,\ Sp_1 = 0.6746,\ Se_2 = 0.8864,\ Sp_2 = 0.6746$ $p = 30\%,\ \varepsilon_1 \leq 0.1007,\ \varepsilon_0 \leq 0.2195$						
$\boldsymbol{n}$	$\varepsilon_1 = 0, \varepsilon_0 = 0$		$\varepsilon_1 = 0.04, \varepsilon_0 = 0.10$		$\varepsilon_1 = 0.08, \varepsilon_0 = 0.20$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.058	0.049	0.045	0.058	0.049	0.045
200	0.050	0.046	0.049	0.050	0.046	0.049
300	0.047	0.046	0.052	0.047	0.046	0.052
400	0.052	0.051	0.048	0.052	0.051	0.048
500	0.048	0.047	0.040	0.048	0.047	0.040
1000	0.049	0.048	0.050	0.049	0.048	0.050
2000	0.046	0.046	0.048	0.046	0.046	0.048

$\kappa_{11} = \kappa_{21} = 0.6$ $Se_1 = 0.43,\ Sp_1 = 0.97,\ Se_2 = 0.43,\ Sp_2 = 0.97$ $p = 5\%,\ \varepsilon_1 \leq 0.2425,\ \varepsilon_0 \leq 0.0291$						
$\boldsymbol{n}$	$\varepsilon_1 = 0, \varepsilon_0 = 0$		$\varepsilon_1 = 0.10, \varepsilon_0 = 0.01$		$\varepsilon_1 = 0.20, \varepsilon_0 = 0.02$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.011	0.002	0.003	0	0	0
200	0.035	0.017	0.019	0.008	0.001	0
300	0.049	0.026	0.024	0.015	0.001	0
400	0.054	0.033	0.040	0.027	0.007	0.006
500	0.054	0.028	0.033	0.023	0.011	0.009
1000	0.047	0.041	0.049	0.044	0.027	0.023
2000	0.055	0.050	0.049	0.046	0.042	0.040

$\kappa_{11} = \kappa_{21} = 0.8$ $Se_1 = 0.8063,\ Sp_1 = 0.9392,\ Se_2 = 0.8063,\ Sp_2 = 0.9392$ $p = 50\%,\ \varepsilon_1 \leq 0.1562,\ \varepsilon_0 \leq 0.0571$						
$\boldsymbol{n}$	$\varepsilon_1 = 0, \varepsilon_0 = 0$		$\varepsilon_1 = 0.07, \varepsilon_0 = 0.02$		$\varepsilon_1 = 0.14, \varepsilon_0 = 0.04$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.033	0.025	0.023	0.019	0.002	0.001
200	0.048	0.045	0.043	0.039	0.011	0.008
300	0.045	0.044	0.036	0.034	0.027	0.024
400	0.053	0.049	0.049	0.047	0.040	0.037
500	0.056	0.055	0.056	0.055	0.037	0.036
1000	0.048	0.048	0.053	0.052	0.043	0.043
2000	0.045	0.045	0.056	0.055	0.051	0.050

Table 3: Type I errors of the hypothesis tests when $L > L'$.

$\kappa_{11} = \kappa_{21} = 0.2$ $Se_1 = 0.4237,\ Sp_1 = 0.8131,\ Se_2 = 0.4237,\ Sp_2 = 0.8131$ $p = 50\%,\ \varepsilon_1 \leq 0.2442,\ \varepsilon_0 \leq 0.1520$						
$\boldsymbol{n}$	$\varepsilon_1 = 0, \varepsilon_0 = 0$		$\varepsilon_1 = 0.02, \varepsilon_0 = 0.10$		$\varepsilon_1 = 0.04, \varepsilon_0 = 0.20$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.047	0.006	0.045	0.013	0.003	0
200	0.048	0.021	0.049	0.027	0.020	0.006
300	0.056	0.037	0.042	0.030	0.030	0.021
400	0.055	0.044	0.052	0.043	0.045	0.034
500	0.058	0.051	0.042	0.037	0.040	0.034
1000	0.046	0.043	0.055	0.052	0.041	0.039
2000	0.046	0.044	0.048	0.048	0.058	0.057

$\kappa_{11} = \kappa_{21} = 0.4$ $Se_1 = 0.7773,\ Sp_1 = 0.7308,\ Se_2 = 0.7773,\ Sp_2 = 0.7308$ $p = 10\%,\ \varepsilon_1 \leq 0.1731,\ \varepsilon_0 \leq 0.1967$						
$\boldsymbol{n}$	$\varepsilon_1 = 0, \varepsilon_0 = 0$		$\varepsilon_1 = 0.08, \varepsilon_0 = 0.09$		$\varepsilon_1 = 0.16, \varepsilon_0 = 0.18$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.022	0.002	0.006	0	0	0
200	0.043	0.026	0.022	0.008	0	0
300	0.055	0.040	0.030	0.018	0.001	0
400	0.049	0.037	0.047	0.038	0.002	0.001
500	0.039	0.032	0.047	0.042	0.002	0.002
1000	0.049	0.047	0.053	0.050	0.014	0.011
2000	0.056	0.054	0.051	0.050	0.030	0.030

$\kappa_{11} = \kappa_{21} = 0.6$ $Se_1 = 0.8864,\ Sp_1 = 0.6746,\ Se_2 = 0.8864,\ Sp_2 = 0.6746$ $p = 30\%,\ \varepsilon_1 \leq 0.1007,\ \varepsilon_0 \leq 0.2195$						
$\boldsymbol{n}$	$\varepsilon_1 = 0, \varepsilon_0 = 0$		$\varepsilon_1 = 0.04, \varepsilon_0 = 0.10$		$\varepsilon_1 = 0.08, \varepsilon_0 = 0.20$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.050	0.034	0.024	0.014	0.001	0
200	0.053	0.048	0.044	0.036	0.009	0.006
300	0.044	0.040	0.058	0.051	0.017	0.015
400	0.053	0.050	0.052	0.048	0.030	0.028
500	0.052	0.050	0.054	0.050	0.033	0.032
1000	0.054	0.054	0.049	0.047	0.051	0.051
2000	0.055	0.054	0.063	0.062	0.056	0.055

$\kappa_{11} = \kappa_{21} = 0.8$ $Se_1 = 0.81,\ Sp_1 = 0.99,\ Se_2 = 0.81,\ Sp_2 = 0.99$ $p = 5\%,\ \varepsilon_1 \leq 0.1539,\ \varepsilon_0 \leq 0.0099$						
$\boldsymbol{n}$	$\varepsilon_1 = 0, \varepsilon_0 = 0$		$\varepsilon_1 = 0.07, \varepsilon_0 = 0.004$		$\varepsilon_1 = 0.14, \varepsilon_0 = 0.008$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.001	0	0	0	0	0
200	0.009	0.008	0.002	0.001	0	0
300	0.016	0.014	0.005	0.002	0	0
400	0.025	0.019	0.010	0.006	0	0
500	0.027	0.024	0.011	0.007	0	0
1000	0.044	0.040	0.037	0.033	0.006	0.003
2000	0.055	0.053	0.043	0.042	0.022	0.019

When $L > L'$ (Table 3), the prevalence and the covariances also have an important effect (and a similar one to the previous situation) upon the type I error of the test $H_0 : \kappa_{12} = \kappa_{22}$. As in the previous situation, the increase in the prevalence implies an increase in the type I error, especially in samples of 100 and 200, although it does not overwhelm the nominal error. The increase in the covariances implies a decrease in the type I error, especially for $n \leq 500$. Therefore, in general terms, when the values of the covariances are high, for a sample size $n \leq 500$ (depending on the disease prevalence) the hypothesis test $H_0 : \kappa_{12} = \kappa_{22}$ is conservative. The prevalence and the covariances have practically no effect upon the type I error when the sample size is very large ($n = 1000 - 2000$). Therefore, in general terms, the type I error of the test $H_0 : \kappa_{12} = \kappa_{22}$ shows very similar behavior to that of the hypothesis test of the comparison of the two average kappa coefficients when $L' > L$ ($H_0 : \kappa_{11} = \kappa_{21}$); i.e. it is a conservative test and starting from a determined sample size its type I error fluctuates around the nominal error without overwhelming it. Regarding the type I error of the test $H_0 : \ln(\kappa_{12}) = \ln(\kappa_{22})$, its behaviour is, in general terms, very similar to that of the test $H_0 : \kappa_{12} = \kappa_{22}$, although for $n = 100 - 200$ its type I error is, as in the case of $L' > L$, somewhat lower than that of the hypothesis test without transformation.

4.2. Powers

In Table 4, we can see some of the results for the power of the hypothesis tests $H_0 : \kappa_{11} = \kappa_{21}$ and $H_0 : \ln(\kappa_{11}) = \ln(\kappa_{21})$, and in Table 5, we can see some of the results for the power of the hypothesis tests $H_0 : \kappa_{12} = \kappa_{22}$ and $H_0 : \ln(\kappa_{12}) = \ln(\kappa_{22})$. In these tables we also indicate the values of the sensitivities, specificities, prevalence and covariances with which the multinomial samples were generated.

When $L' > L$ (Table 4), the disease prevalence has an important effect on the powers of the tests $H_0 : \kappa_{11} = \kappa_{21}$ and $H_0 : \ln(\kappa_{11}) = \ln(\kappa_{21})$. For the same sample size, the power of each hypothesis test rises with an increase in the prevalence. Regarding the covariances between the two BDTs, the power also rises with increase in the covariances, although its effect is, in general terms, less important than in the case of prevalence. Consequently, based on the prevalence we can reach the following general conclusions:

1. For a prevalence equal to 5% it is necessary to have very large sample size ($n \geq 1000$) so that the power is high (above 80%). If the prevalence is equal to 10%, with a sample size $n \geq 200$ high power is obtained (above 80%, depending on the covariances).
2. If the prevalence is high, p equal to 30% or 50%, with a sample size $n \geq 200$ the powers of both hypothesis tests are very high (higher than 80% or 90%, depending on the covariances).

Table 4: Powers of the hypothesis tests when $L' > L$.

$\kappa_{11} = 0.4,\ \kappa_{21} = 0.2$ $Se_1 = 0.8209,\ Sp_1 = 0.8670,\ Se_2 = 0.7773,\ Sp_2 = 0.7308$ $p = 10\%,\ \varepsilon_1 \leq 0.1392,\ \varepsilon_0 \leq 0.0972$						
$\boldsymbol{n}$	$\varepsilon_1 = 0,\ \varepsilon_0 = 0$		$\varepsilon_1 = 0.06,\ \varepsilon_0 = 0.04$		$\varepsilon_1 = 0.12,\ \varepsilon_0 = 0.08$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.498	0.452	0.613	0.593	0.767	0.755
200	0.831	0.837	0.927	0.935	0.995	0.996
300	0.937	0.941	0.987	0.988	1	1
400	0.986	0.987	1	1	1	1
500	0.990	0.991	1	1	1	1
1000	1	1	1	1	1	1
2000	1	1	1	1	1	1

$\kappa_{11} = 0.6,\ \kappa_{21} = 0.4$ $Se_1 = 0.8495,\ Sp_1 = 0.8375,\ Se_2 = 0.8864,\ Sp_2 = 0.6746$ $p = 30\%,\ \varepsilon_1 \leq 0.0965,\ \varepsilon_0 \leq 0.1096$						
$\boldsymbol{n}$	$\varepsilon_1 = 0,\ \varepsilon_0 = 0$		$\varepsilon_1 = 0.04,\ \varepsilon_0 = 0.04$		$\varepsilon_1 = 0.08,\ \varepsilon_0 = 0.08$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.570	0.554	0.692	0.677	0.820	0.812
200	0.845	0.844	0.917	0.916	0.985	0.986
300	0.939	0.939	0.982	0.983	0.998	0.998
400	0.984	0.984	0.997	0.997	1	1
500	0.992	0.992	0.998	0.998	1	1
1000	1	1	1	1	1	1
2000	1	1	1	1	1	1

$\kappa_{11} = 0.6,\ \kappa_{21} = 0.2$ $Se_1 = 0.8991,\ Sp_1 = 0.7458,\ Se_2 = 0.8131,\ Sp_2 = 0.4237$ $p = 50\%,\ \varepsilon_1 \leq 0.0820,\ \varepsilon_0 \leq 0.1076$						
$\boldsymbol{n}$	$\varepsilon_1 = 0,\ \varepsilon_0 = 0$		$\varepsilon_1 = 0.06,\ \varepsilon_0 = 0.01$		$\varepsilon_1 = 0.12,\ \varepsilon_0 = 0.02$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.954	0.907	0.973	0.949	0.982	0.961
200	0.998	0.998	0.998	0.998	0.999	0.999
300	1	1	1	1	1	1
400	1	1	1	1	1	1
500	1	1	1	1	1	1
1000	1	1	1	1	1	1
2000	1	1	1	1	1	1

$\kappa_{11} = 0.8,\ \kappa_{21} = 0.6$ $Se_1 = 0.81,\ Sp_1 = 0.99,\ Se_2 = 0.62,\ Sp_2 = 0.98$ $p = 5\%,\ \varepsilon_1 \leq 0.1178,\ \varepsilon_0 \leq 0.0098$						
$\boldsymbol{n}$	$\varepsilon_1 = 0,\ \varepsilon_0 = 0$		$\varepsilon_1 = 0.05,\ \varepsilon_0 = 0.004$		$\varepsilon_1 = 0.10,\ \varepsilon_0 = 0.008$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.010	0.002	0.005	0	0.001	0
200	0.097	0.062	0.100	0.055	0.085	0.049
300	0.252	0.207	0.260	0.208	0.272	0.210
400	0.365	0.323	0.396	0.364	0.466	0.404
500	0.483	0.442	0.520	0.483	0.615	0.586
1000	0.735	0.721	0.801	0.797	0.842	0.842
2000	0.890	0.890	0.890	0.888	0.895	0.895

Table 5: Powers of the hypothesis tests when $L > L'$.

$\kappa_{11} = 0.4$, $\kappa_{21} = 0.2$ $Se_1 = 0.7021$, $Sp_1 = 0.6817$, $Se_2 = 0.3019$, $Sp_2 = 0.9030$ $p = 30\%$, $\varepsilon_1 \leq 0.0900$, $\varepsilon_0 \leq 0.0661$						
$\boldsymbol{n}$	$\varepsilon_1 = 0$, $\varepsilon_0 = 0$		$\varepsilon_1 = 0.04$, $\varepsilon_0 = 0.03$		$\varepsilon_1 = 0.08$, $\varepsilon_0 = 0.06$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.410	0.238	0.416	0.252	0.433	0.278
200	0.630	0.587	0.733	0.693	0.784	0.749
300	0.790	0.773	0.862	0.851	0.931	0.927
400	0.878	0.876	0.938	0.936	0.978	0.977
500	0.941	0.940	0.970	0.969	0.991	0.991
1000	0.998	0.998	1	1	1	1
2000	1	1	1	1	1	1

$\kappa_{11} = 0.6$, $\kappa_{21} = 0.4$ $Se_1 = 0.8624$, $Sp_1 = 0.6816$, $Se_2 = 0.8112$, $Sp_2 = 0.5293$ $p = 50\%$, $\varepsilon_1 \leq 0.1116$, $\varepsilon_0 \leq 0.1686$						
$\boldsymbol{n}$	$\varepsilon_1 = 0$, $\varepsilon_0 = 0$		$\varepsilon_1 = 0.05$, $\varepsilon_0 = 0.07$		$\varepsilon_1 = 0.10$, $\varepsilon_0 = 0.14$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.434	0.347	0.567	0.473	0.731	0.623
200	0.680	0.650	0.840	0.820	0.987	0.984
300	0.825	0.813	0.948	0.945	0.999	0.999
400	0.901	0.899	0.981	0.980	1	1
500	0.956	0.952	0.996	0.995	1	1
1000	1	0.999	1	1	1	1
2000	1	1	1	1	1	1

$\kappa_{11} = 0.6$, $\kappa_{21} = 0.2$ $Se_1 = 0.8209$, $Sp_1 = 0.8670$, $Se_2 = 0.2091$, $Sp_2 = 0.9715$ $p = 10\%$, $\varepsilon_1 \leq 0.0374$, $\varepsilon_0 \leq 0.0247$						
$\boldsymbol{n}$	$\varepsilon_1 = 0$, $\varepsilon_0 = 0$		$\varepsilon_1 = 0.015$, $\varepsilon_0 = 0.01$		$\varepsilon_1 = 0.03$, $\varepsilon_0 = 0.02$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.376	0.135	0.359	0.146	0.411	0.164
200	0.805	0.683	0.814	0.693	0.838	0.720
300	0.945	0.914	0.965	0.928	0.874	0.947
400	1	0.978	0.993	0.972	0.996	0.990
500	1	1	1	1	1	1
1000	1	1	1	1	1	1
2000	1	1	1	1	1	1

$\kappa_{11} = 0.8$, $\kappa_{21} = 0.6$ $Se_1 = 0.9528$, $Sp_1 = 0.9598$, $Se_2 = 0.62$, $Sp_2 = 0.98$ $p = 5\%$, $\varepsilon_1 \leq 0.0292$, $\varepsilon_0 \leq 0.0191$						
$\boldsymbol{n}$	$\varepsilon_1 = 0$, $\varepsilon_0 = 0$		$\varepsilon_1 = 0.01$, $\varepsilon_0 = 0.07$		$\varepsilon_1 = 0.02$, $\varepsilon_0 = 0.14$	
	Test 1	**Test 2**	**Test 1**	**Test 2**	**Test 1**	**Test 2**
100	0.017	0.005	0.024	0.004	0.019	0.007
200	0.112	0.067	0.109	0.057	0.123	0.067
300	0.233	0.189	0.229	0.164	0.243	0.191
400	0.391	0.331	0.401	0.325	0.368	0.308
500	0.483	0.440	0.480	0.428	0.510	0.468
1000	0.796	0.777	0.835	0.822	0.839	0.826
2000	0.953	0.953	0.944	0.944	0.951	0.951

Finally, in general terms, the test $H_0 : \kappa_{11} = \kappa_{21}$ is more powerful than the test $H_0 : \ln(\kappa_{11}) = \ln(\kappa_{21})$, especially when $n \leq 200$, since its type I error is slightly greater (without overwhelming the nominal error).

When $L > L'$ (Table 5), the powers of the hypothesis tests $H_0 : \kappa_{12} = \kappa_{22}$ and $H_0 : \ln(\kappa_{12}) = \ln(\kappa_{22})$ show very similar behaviour to that of the previous case ($L' > L$). The disease prevalence and the covariances have a very similar effect, and the conclusions about the powers are also very similar, although when the prevalence is 10% it is necessary to have a slightly larger sample size ($n \geq 200 - 300$) so that the power is high (above 80%). Finally, and as in the previous case, the test $H_0 : \kappa_{12} = \kappa_{22}$ is more powerful than the test $H_0 : \ln(\kappa_{12}) = \ln(\kappa_{22})$, especially when $n \leq 200$, since its type I error is also slightly greater (without overwhelming the nominal error).

5. EXTENSION TO MORE THAN TWO BDTS

Let us consider J BDTs ($J \geq 3$) and a GS that are applied to all of the n individuals in a random sample. When $L' > L$, the expression of the weighted kappa coefficient for the j-th BDT is

$$\kappa_{j1} = \begin{cases} \frac{2\kappa_j(0)\kappa_j(1)}{\kappa_j(0)-\kappa_j(1)} \ln\left[\frac{\kappa_j(0)+\kappa_j(1)}{2\kappa_j(1)}\right], & p \neq Q_j \\ Se_j + Sp_j - 1, & p = Q_j \end{cases}$$

and when $L > L'$ its expression is

$$\kappa_{j2} = \begin{cases} \frac{2\kappa_j(0)\kappa_j(1)}{\kappa_j(0)-\kappa_j(1)} \ln\left[\frac{2\kappa_j(0)}{\kappa_j(0)+\kappa_j(1)}\right], & p \neq Q_j \\ Se_j + Sp_j - 1, & p = Q_j, \end{cases}$$

with $\kappa_j(0) = \frac{Sp_j-(1-Q_j)}{Q_j}$, $\kappa_j(1) = \frac{Se_j-Q_j}{1-Q_j}$ and $Q_j = pSe_j + q(1 - Sp_j)$, and where $p = \sum_{i_1,\ldots,i_J=0}^{1} p_{i_1,\ldots,i_J}$ is the disease prevalence and $q = 1-p = \sum_{i_1,\ldots,i_J=0}^{1} q_{i_1,\ldots,i_J}$. The sensitivity and the specificity of the j-th BDT are written as

$$Se_j = \frac{\sum_{\substack{i_1,\ldots,i_J=0 \\ i_j=1}}^{1} p_{i_1,\ldots,i_J}}{\sum_{i_1,\ldots,i_J=0}^{1} p_{i_1,\ldots,i_J}}$$

and

$$Sp_j = \frac{\sum_{\substack{i_1,\ldots,i_J=0 \\ i_j=0}}^{1} q_{i_1,\ldots,i_J}}{\sum_{i_1,\ldots,i_J=0}^{1} q_{i_1,\ldots,i_J}},$$

respectively. Replacing these expressions with those of each average kappa coefficient, then

$$\kappa_{j1} = \begin{cases} \frac{2}{a_1 - a_2} \times \ln\left[\frac{b_1+1}{2}\right], & p \neq Q_j \\ Se_j + Sp_j - 1, & p = Q_j \end{cases}$$

and

$$\kappa_{j2} = \begin{cases} \frac{2}{a_1 - a_2} \times \ln\left[\frac{2}{b_2+1}\right], & p \neq Q_j \\ Se_j + Sp_j - 1, & p = Q_j, \end{cases}$$

where

$$a_1 = \frac{p - \sum\limits_{\substack{i_1,\ldots,i_J=0 \\ i_j=1}}^{1} p_{i_1,\ldots,i_J} + \sum\limits_{\substack{i_1,\ldots,i_J=0 \\ i_j=0}}^{1} q_{i_1,\ldots,i_J}}{\frac{\sum\limits_{\substack{i_1,\ldots,i_J=0 \\ i_j=1}}^{1} p_{i_1,\ldots,i_J}}{\sum\limits_{i_1,\ldots,i_J=0}^{1} p_{i_1,\ldots,i_J}} - q - \sum\limits_{\substack{i_1,\ldots,i_J=0 \\ i_j=1}}^{1} p_{i_1,\ldots,i_J} + \sum\limits_{\substack{i_1,\ldots,i_J=0 \\ i_j=0}}^{1} q_{i_1,\ldots,i_J}},$$

$$a_2 = \frac{q + \sum\limits_{\substack{i_1,\ldots,i_J=0 \\ i_j=1}}^{1} p_{i_1,\ldots,i_J} - \sum\limits_{\substack{i_1,\ldots,i_J=0 \\ i_j=0}}^{1} q_{i_1,\ldots,i_J}}{\frac{\sum\limits_{\substack{i_1,\ldots,i_J=0 \\ i_j=0}}^{1} q_{i_1,\ldots,i_J}}{\sum\limits_{i_1,\ldots,i_J=0}^{1} q_{i_1,\ldots,i_J}} - p + \sum\limits_{\substack{i_1,\ldots,i_J=0 \\ i_j=1}}^{1} p_{i_1,\ldots,i_J} - \sum\limits_{\substack{i_1,\ldots,i_J=0 \\ i_j=0}}^{1} q_{i_1,\ldots,i_J}},$$

$b_1 = \frac{a_1}{a_2}$ and $b_2 = \frac{1}{b_1}$. As the maximum likelihood estimators of the probabilities $p_{i_1,\ldots,i_J}$ and $q_{i_1,\ldots,i_J}$ are $\hat{p}_{i_1,\ldots,i_J} = s_{i_1,\ldots,i_J}/n$ and $\hat{q}_{i_1,\ldots,i_J} = r_{i_1,\ldots,i_J}/n$, with $i_1, \ldots, i_J = 0, 1$, the estimator of each average kappa coefficient is obtained replacing in the expressions of κ_{j1} and κ_{j2} each parameter $p_{i_1,\ldots,i_J}$ and $q_{i_1,\ldots,i_J}$ with its corresponding estimator. Let $\boldsymbol{\kappa}_i = (\kappa_{1i}, \kappa_{2i}, \ldots, \kappa_{Ji})^T$ be the vector of average kappa coefficients and $\hat{\boldsymbol{\kappa}}_i = (\hat{\kappa}_{1i}, \hat{\kappa}_{2i}, \ldots, \hat{\kappa}_{Ji})^T$ its estimator, where $i = 1$ when $L' > L$ and $i = 2$ when $L > L'$. Applying the delta method, the asymptotic variances-covariances matrix of the vector $\hat{\boldsymbol{\kappa}}_i$ is $\sum_{\hat{\boldsymbol{\kappa}}_i} = \left(\frac{\partial \boldsymbol{\kappa}_i}{\partial \boldsymbol{\pi}}\right) \sum_{\hat{\boldsymbol{\pi}}} \left(\frac{\partial \boldsymbol{\kappa}_i}{\partial \boldsymbol{\pi}}\right)^T$, where $\boldsymbol{\pi}$ is the vector of probabilities. Performing algebraic operations and replacing in this expression each parameter with its estimator, the estimated asymptotic variances-covariances matrix $\hat{\sum}_{\hat{\boldsymbol{\kappa}}_i}$ is obtained. The global hypothesis test to contrast the equality of the J average kappa coefficients is $H_0 : \kappa_{1i} = \kappa_{2i} = \ldots = \kappa_{Ji}$ vs H_1 : at least one equality is not true. This hypothesis test is equivalent to $H_0 : \boldsymbol{\varphi}\boldsymbol{\kappa}_i = \mathbf{0}$ vs $H_1 : \boldsymbol{\varphi}\boldsymbol{\kappa}_i \neq \mathbf{0}$, where $\boldsymbol{\varphi}$ is a complete range matrix whose dimension is $(J-1) \times J$. For example, for three BDTs the matrix $\boldsymbol{\varphi}$ is

$$\boldsymbol{\varphi} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

Applying the multivariate central limit theorem it is verified that

$$\sqrt{n}\left(\hat{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_i\right) \xrightarrow[n\to\infty]{} N_{J-1}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\kappa}_i}\right),$$

so that the statistic $Q^2 = \hat{\boldsymbol{\kappa}}_i{}^T \boldsymbol{\varphi}^T \left(\boldsymbol{\varphi} \hat{\sum}_{\hat{\boldsymbol{\kappa}}_i} \boldsymbol{\varphi}^T \right)^{-1} \boldsymbol{\varphi} \hat{\boldsymbol{\kappa}}_i$ is distributed according to a distribution T^2 of Hotelling sized $J-1$ and n degrees of freedom, where $J-1$ is the dimension of vector $\boldsymbol{\varphi}\hat{\boldsymbol{\kappa}}_i$. For a large n, the statistic Q^2 is distributed according to a chi-squared central distribution with $J-1$ degrees of freedom when the null hypothesis is true, i.e. $Q^2 = \hat{\boldsymbol{\kappa}}_i{}^T \boldsymbol{\varphi}^T \left(\boldsymbol{\varphi} \hat{\sum}_{\hat{\boldsymbol{\kappa}}_i} \boldsymbol{\varphi}^T \right)^{-1} \boldsymbol{\varphi} \hat{\boldsymbol{\kappa}}_i \xrightarrow[n\to\infty]{} \chi^2_{J-1}$.

The procedure to solve the hypothesis test would be very similar to that used by Roldán-Nofuentes *et al.* [9] to simultaneously compare the weighted kappa coefficients of multiple BDTs: 1) solve the global test to an error of α; 2) if the global test is not significant at that error rate, then the homogeneity of the J average kappa coefficients is not rejected, and if the test is significant then the investigation into the causes of the significance is carried out comparing the pairs of average kappa coefficients using the results in Section 3 and penalizing the level of significance through some method of multiple comparisons, for example Bonferroni [10], Holm [11] or Hochberg [12].

Finally, as in the case of two BDTs, the comparison of multiple average kappa coefficients can be made using logarithmic transformation, and the procedure is similar to that used in the case without transformation.

6. THE "CAKCTBT" PROGRAM

The "cakctbt" program (Comparison of Average Kappa Coefficients of Two Binary Tests) is a program written in R that solves the hypothesis tests to contrast the equality of the average kappa coefficients of two BDTs, i.e. $H_0 : \kappa_{11} = \kappa_{21}$ and $H_0 : \kappa_{12} = \kappa_{22}$. This program runs with the command

$$\text{cakctbt}\,(s_{11}, s_{10}, s_{01}, s_{00}, r_{11}, r_{10}, r_{01}, r_{00})$$

when $\alpha = 5\%$, and with the command

$$\text{cakctbt}\,(s_{11}, s_{10}, s_{01}, s_{00}, r_{11}, r_{10}, r_{01}, r_{00}, \alpha)$$

when $\alpha \neq 5\%$. The program provides the estimation of each average kappa coefficient and its respective standard error, the value of the contrast statistic and the p-value of each hypothesis test. It also provides the confidence intervals for the difference of the average kappa coefficients in each situation ($L' > L$ and $L > L'$). The results obtained when running the program are kept in a file called "Results_cakctbt.txt" in the same folder from where the program is run. The program is available for free at URL:

"`http://www.ugr.es/~bioest/software/cmd.php?seccion=mdb`".

7. EXAMPLE

The results in Section 3 were applied to the study of Weiner *et al.* [13] about the diagnosis of coronary disease, which is a classic example when comparing the parameters of two BDTs subject to a paired design. In Table 6 (Observed frequencies), we can see the results when applying two BDTs, a cardiac stress test and the individual's clinical history in relation to coronary disease, and the GS (coronary arteriography) to a sample of 871 individuals, and where the variable T_1 models the result of the stress test, T_2 models the result of the individual's clinical history and the variable D models the result of the coronary angiography.

Table 6: Data of the study of Weiner *et al.* and results.

Observed frequencies					
	$T_1 = 1$		$T_1 = 0$		Total
	$T_2 = 1$	$T_2 = 0$	$T_2 = 1$	$T_2 = 0$	
$D = 1$	473	29	81	25	608
$D = 0$	22	46	44	151	263
Total	495	75	125	176	871

Results
$L' > L$
$\hat{\kappa}_{11} = 0.574$, $\hat{\kappa}_{21} = 0.658$ $\hat{Var}(\hat{\kappa}_{11}) = 0.031820$, $\hat{Var}(\hat{\kappa}_{21}) = 0.029746$ $\hat{Cov}(\hat{\kappa}_{11}, \hat{\kappa}_{21}) = 0.000112$
$H_0 : \kappa_{11} = \kappa_{21}$ vs $H_1 : \kappa_{11} \neq \kappa_{21}$ $z = 2.06$, p-value $= 0.039$ 95% CI for $\kappa_{21} - \kappa_{11}$: (0.0041 ; 0.1644)
$L > L'$
$\hat{\kappa}_{12} = 0.519$, $\hat{\kappa}_{22} = 0.680$ $\hat{Var}(\hat{\kappa}_{12}) = 0.031303$, $\hat{Var}(\hat{\kappa}_{22}) = 0.029260$ $\hat{Cov}(\hat{\kappa}_{11}, \hat{\kappa}_{21}) = 0.000229$
$H_0 : \kappa_{12} = \kappa_{22}$ vs $H_1 : \kappa_{12} \neq \kappa_{22}$ $z = 4.33$, p-value $= 1.46 \times 10^{-5}$ 95% CI for $\kappa_{22} - \kappa_{12}$: (0.0881 ; 0.2336)

In Table 6 (Results), we can see the estimations of the parameters, the results of the hypothesis tests ($\alpha = 5\%$) and the confidence intervals to 95%. Based on these results, if the clinical laboratory researcher is more concerned about the false positives than the false negatives ($L' > L$), then the equality of the average kappa coefficients is rejected, and it holds that the average kappa coefficient of the clinical history (which has a "good" value in terms of point estimation) is signifi-

cantly larger than that of the stress test (which has a "moderate" value in terms of point estimation). Therefore, the average beyond-chance agreement between the clinical history and the angiography is, with a confidence of 95%, a value between 0.0041 and 0.1644 greater than the average beyond-chance agreement between the stress tests and the angiography. Similar conclusions are obtained if the clinical laboratory researcher is more concerned about the false negatives than the false positives ($L > L'$). In this situation, the average beyond-chance agreement between the clinical history and the angiography, with a confidence of 95%, is a value between 0.0881 and 0.2336 higher than the average beyond-chance agreement between the stress tests and the angiography.

8. DISCUSSION

The comparison of the performance of two BDTs in relation to a GS can be made through a paired design or an unpaired one. Paired design consists of applying the two BDTs to all of the individuals in a simple, whereas in unpaired design each individual is only tested with one of the two BDTs. Paired design is used more in practice and has more advantages than unpaired design [14]. Paired design was chosen to develop the method proposed in this article.

In clinical practice, when we consider the losses in an erroneous classification with two BDTs, the appropriate parameters to compare the two BDTs are weighted kappa coefficients. In this situation, it is necessary to assume a value for the weighting index c and solve the test $H_0 : \kappa_1(c) = \kappa_2(c)$ vs $H_1 : \kappa_1(c) \neq \kappa_2(c)$ applying the Bloch method [4]. The value of the weighting index c is set by the clinical laboratory researcher based on his or her knowledge about the problem in question. If the clinical laboratory researcher does not have enough knowledge to allow them to allocate a value to the weighting index c, the comparison of the performance of the two (or more) BDTs can be made through the average kappa coefficients, which are measures of the beyond-chance agreement between each BDT and the GS and do not depend on the weighting index c. Therefore, if the clinical laboratory researcher can assume a value of the weighting index c, then compare the weighted kappa coefficients of the two BDTs applying the Bloch method [4]. In the opposite case, compare the weighted kappa coefficients κ_{i1} if there is a greater concern about the false positives than about the false negatives, or compare the weighted kappa coefficients κ_{i2} if there is a greater concern about the false negatives than the false positives.

In this article, we have studied the comparison of the average kappa coefficients of two (and more) BDTs when the clinical laboratory researcher considers that loss associated with the false positives is greater than that associated with the false negatives ($L' > L$), and when the clinical laboratory researcher considers the opposite ($L > L'$). The hypothesis tests studied are asymptotic and the

simulation experiments carried out have demonstrated that the type I errors do not overwhelm the nominal error of 5%. Regarding the power of each hypothesis test, this increases with the prevalence, and so when the prevalence is small (e.g. 5%) it is necessary to have a very large sample size ($n \geq 1000$) so that the power is high (above 80%); whereas with a large prevalence (e.g. 30% or 50%), with a sample size $n \geq 200$ a high power is obtained.

In the expressions of the statistics deduced to solve the hypothesis tests, the variances-covariances have been estimated applying the delta method. An alternative method is to estimate these variances-covariances through bootstrap. Simulation experiments (similar to those in Section 4) have shown that there is no important difference in terms of the type I error and the power between both methods of estimation of the variances-covariances.

The results were extended to the case of more than two BDTs, finding that the solution to the hypothesis test is also asymptotic and a method based on multiple comparisons is proposed to solve the problem. This method is very similar to that used in the analysis of the variance. Firstly, the global test is solved to an error of α and if the test is significant then the causes of the significance are investigated making paired comparisons and applying a multiple comparison method. For our problem, we have chosen the Bonferroni, Holm or Hochberg methods, which are very easy to apply and have been used in the field of BDTs [15,16].

The method that we have proposed requires knowledge of the disease status of all of the individuals in a sample through the application of the GS. If the disease status of any individual is unknown, leading to the problem known as partial disease verification, the method proposed cannot be applied. If the verification process with the GS only depends on the results of the BDTs, a solution to this problem could be obtained following a method similar to that used by Roldán-Nofuentes and Luna del Castillo [17] and Roldán-Nofuentes *et al.* [18].

If case-control sampling is being used, the method that we have proposed cannot be used either as it is necessary to know the disease prevalence. An extension of the study of Roldán-Nofuentes and Amro [19] to the situation of two BDTs may be a solution to this problem.

ACKNOWLEDGMENTS

This research was supported by the Spanish Ministry of Economy, Grant Number MTM2016-76938-P. We thank the two Referees, Editor Associate and Editor of REVSTAT – Statistical Journal for their helpful comments that improved the quality of the paper.

REFERENCES

[1] Bloch, D.A. and Kraemer, H.C. (1989). 2×2 Kappa coefficients: measures of agreement or association, *Biometrics*, **45**, 269–287.

[2] Kraemer, H.C. (1992). *Evaluating medical tests. Objective and quantitative guidelines*, Sage Publications, Newbury Park.

[3] Kraemer, H.C.; Periyakoil, V.S. and Noda, A. (1992). Kappa coefficients in medical research, *Statistics in Medicine*, **21**, 2109–2129.

[4] Bloch, D.A. (1997). Comparing two diagnostic tests against the same "gold standard" in the same sample, *Biometrics*, **53**, 73–85.

[5] Roldán-Nofuentes, J.A. and Olvera-Porcel, C. (2015). Average kappa coefficient: a new measure to assess a binary test considering the losses associated with an erroneous classification, *Journal of Statistical Computation and Simulation*, **85**, 1601–1620.

[6] Roldán-Nofuentes, J.A.; Luna del Castillo, J.D. and Montero-Alonso, M.A. (2009). Confidence intervals of weighted kappa coefficient of a binary diagnostic test, *Communications in Statistics – Simulation and Computation*, **38**, 1562–1578.

[7] Vacek, P. (1985). The effect of conditional dependence on the evaluation of diagnostic tests, *Biometrics*, **41**, 959–968.

[8] Cicchetti, D.V. (2001). The precision of reliability and validity estimates revisited: distinguishing between clinical and statistical significance of sample size requirements, *Journal of Clinical and Experimental Neuropsychology*, **23**, 695–700.

[9] Roldán-Nofuentes, J.A. and Luna del Castillo, J.D. (2010). Comparison of weighted kappa coefficients of multiple binary diagnostic tests done on the same subjects, *Statistics in Medicine*, **29**, 2149–2165.

[10] Bonferroni, C.E. (1936). Teoria statistica delle classi e calcolo delle probabilitá, *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, **8**, 3–62.

[11] Holm, S. (1979). A simple sequential rejective multiple testing procedure, *Scandinavian Journal of Statistics*, **6**, 65–70.

[12] Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance, *Biometrika*, **75**, 800–802.

[13] Weiner, D.A.; Ryan, T.J.; McCabe, C.H.; Kennedy, J.W.; Schloss, M.; Tristani, F.; Chaitman, B.R. and Fisher, L.D. (1979). Exercise stress testing. Correlations among history of angina, ST-segment response and prevalence of coronary-artery disease in the coronary artery surgery study (CASS), *The New England Journal of Medicine*, **31**, 230–235.

[14] Pepe, M.S. (2003). *The statistical evaluation of medical tests for classification and prediction*, Oxford University Press, New York.

[15] Roldán Nofuentes, J.A.; Luna del Castillo, J.D. and Montero Alonso, M.A. (2012). Global hypothesis test to simultaneously compare the predictive values of two binary diagnostic tests, *Computational Statistics and Data Analysis, Special issue "Computational Statistics for Clinical Research"*, **56**, 1161–1173.

[16] Marín-Jiménez, E. and Roldán-Nofuentes, J.A. (2015). Comparison of the predictive values of multiple binary diagnostic tests in the presence of ignorable missing data, *Revstat*, **15**, 45–64.

[17] Roldán-Nofuentes, J.A. and Luna del Castillo, J.D. (2006). Comparing two binary diagnostic tests in the presence of verification bias, *Computational Statistics and Data Analysis*, **50**, 1551–1564.

[18] Roldán-Nofuentes, J.A.; Marín-Jiménez, E. and Luna del Castillo, J.D. (2014). Asymptotic hypothesis test to simultaneously compare the weighted kappa coefficients of multiple binary diagnostic tests in the presence of ignorable missing data, *Journal of Statistical Computation and Simulation*, **84**, 273–289.

[19] Roldán-Nofuentes, J.A. and Amro, R. (2017). Approximate confidence intervals for the weighted kappa coefficient of a binary diagnostic test subject to a case-control design, *Journal of Statistical Computation and Simulation*, **87**, 530–545.

REVSTAT – Statistical Journal
Volume 16, Number 4, October 2018, 429–443

PARAMETER ESTIMATION FOR THE LOG-LOGISTIC DISTRIBUTION BASED ON ORDER STATISTICS

Authors: Mohammad Ahsanullah
– Department of Management Sciences, Rider University, New Jersey, USA
ahsan@rider.edu

Ayman Alzaatreh
– Department of Mathematics and Statistics, American University of Sharjah, Sharjah, UAE
aalzaatreh@aus.edu

Received: April 2016 Revised: September 2016 Accepted: October 2016

Abstract:

- In this paper, we discuss the moments and product moments of the order statistics in a sample of size n drawn from the log-logistic distribution. We provide more compact forms for the mean, variance and covariance of order statistics. Parameter estimation for the log-logistic distribution based on order statistics is studied. In particular, best linear unbiased estimators (BLUEs) for the location and scale parameters for the log-logistic distribution with known shape parameter are studied. Hill estimator is proposed for estimating the shape parameter.

Key-Words:

- *log-logistic distribution; moments; order statistics; best linear unbiased estimators; recurrence relations; Hill estimator.*

AMS Subject Classification:

- 62E10, 62F10.

1. INTRODUCTION

The probability density function (pdf) of the log-logistic distribution with unit scale parameter is given by

$$f(x) = \frac{\alpha\, x^{\alpha-1}}{(1+x^{\alpha})^2}, \quad x \geq 0, \tag{1.1}$$

where α is a positive real number. A random variable X that follows the density function in (1.1) is denoted as $X \sim$ log-logistic(α). The cumulative distribution (cdf) and quantile functions of the log-logistic distribution, respectively, are

$$F(x) = \frac{x^{\alpha}}{1+x^{\alpha}}, \quad x \geq 0. \tag{1.2}$$

and

$$F^{-1}(x) = \left(\frac{x}{1-x}\right)^{1/\alpha}, \quad 0 < x < 1. \tag{1.3}$$

The k-th moments of the log-logistic distribution in (1.1) can be easily computed as

$$\mu'_k = B\left(1-\frac{k}{\alpha}, 1+\frac{k}{\alpha}\right), \tag{1.4}$$

where $B(.,.)$ is the beta function.
Note that the k-th moment exists iff $\alpha > k$. A more compact form of (1.4) can be derived using the fact that $\Gamma(z)\,\Gamma(1-z) = \pi\, \csc(\pi\, z)$ (Abramowitz and Stegun, 1964) as follows

$$\mu'_k = \Gamma(1-k/\alpha)\,\Gamma(1+k/\alpha) = \frac{k\pi}{\alpha}\, \csc\frac{k\pi}{\alpha}, \quad \alpha > k. \tag{1.5}$$

Therefore,

$$E(X) = (\pi/\alpha)\csc(\pi/\alpha) \text{ and } Var(X) = (\pi/\alpha)\{2\csc(2\pi/\alpha) - (\pi/\alpha)\csc^2(\pi/\alpha)\}.$$

The log-logistic distribution is a well-known distribution and it is used in different fields of study such as survival analysis, hydrology and economy. For some applications of the log-logistic distribution we refer the reader to Shoukri *et al.* [23], Bennett [10], Collett [11] and Ashkar and Mahdi [7]. It is also known that the log-logistic distribution provides good approximation to the normal and the log-normal distributions. The log-logistic distribution has been studied by many researchers such as Shah and Dave [22], Tadikamalla and Johnson [24], Ragab and Green [21], Voorn [25] and Ali and Khan [4]. Ragab and Green [21] studied some properties of the order statistics from the log-logistic distribution. Ali and Khan [4] obtained several recurrence relations for the moments of order

statistics. Voorn [25] characterized the log-logistic distribution based on extreme related stability with random sample size. In this paper, we discuss the moments of order statistics for the log-logistic distribution. We review some known results and provide a more compact expression for calculating the covariance between two order statistics. Also, we discuss the parameter estimation of the log-logistic distribution based on order statistics.

2. SOME RESULTS FOR THE MOMENTS OF ORDER STATISTICS

Let $X_1, X_2, ..., X_n$ be n independent copies of a random variable X that follows log-logistic(α). Let $\mathrm{X}_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ be the corresponding order statistics. Then from (1.1) and (1.2), the pdf of the r^{th} order statistics is given by

$$f_{r:n}(x) = C_{r:n}\, \frac{\alpha\, x^{\alpha r-1}}{(1+x^{\alpha})^{n+1}}, \quad x \geq 0, \tag{2.1}$$

where $C_{r:n} = \frac{n!}{(r-1)!(n-r)!}$.
The k^{th} moments of $X_{r:n}$ can be easily derived from (2.1) as

$$\alpha^{(k)}_{r:n} = C_{r:n}\, B\left(n-r+1-\frac{k}{\alpha}, r+\frac{k}{\alpha}\right), \quad \alpha > k, \tag{2.2}$$

Similarly as in (1.5), one can show that

$$\alpha^{(k)}_{r:n} = \frac{(-1)^r \pi \,\csc\frac{k\pi}{\alpha}}{(r-1)!(n-r)!} \prod_{i=1}^{n}\left(i-r-\frac{k}{\alpha}\right), \quad \alpha > k. \tag{2.3}$$

Note that when $r = n = 1$, $\alpha^{(k)}_{1:1} = B\left(1-\frac{k}{\alpha}, 1+\frac{k}{\alpha}\right)$ which agrees with (1.4). From (2.2), the first and second moments of $X_{r:n}$ are, respectively, given by

$$\alpha^{(1)}_{r:n} = \frac{(-1)^r \pi \,\csc\frac{\pi}{\alpha}}{(r-1)!(n-r)!} \prod_{i=1}^{n}\left(i-r-\frac{1}{\alpha}\right), \quad \alpha > 1, \tag{2.4}$$

and

$$\alpha^{(2)}_{r:n} = \frac{(-1)^r \pi \,\csc\frac{2\pi}{\alpha}}{(r-1)!(n-r)!} \prod_{i=1}^{n}\left(i-r-\frac{2}{\alpha}\right), \quad \alpha > 2. \tag{2.5}$$

It is interesting to note that (2.3) can be used easily to derive several recurrence relations for the moments of order statistics. Some of these recurrence relations already exist in the literature. Below, we provide some of these recurrence relations.

I. From (2.3), we can write

$$\begin{aligned}\alpha_{r:n}^{(k)} &= \frac{-1}{r-1}\,\frac{(-1)^{r-1}\pi\,\csc\frac{k\pi}{\alpha}}{(r-2)!(n-r)!}\prod_{i=0}^{n-1}\left(i-(r-1)-\frac{k}{\alpha}\right)\\ &= \frac{r-1+k/\alpha}{r-1}\,\frac{(-1)^{r-1}\pi\,\csc\frac{k\pi}{\alpha}}{(r-2)!(n-r)!}\prod_{i=1}^{n-1}\left(i-(r-1)-\frac{k}{\alpha}\right)\\ &= \left[1+\frac{k}{\alpha(r-1)}\right]\alpha_{r-1:n-1}^{(k)},\quad 2\le r\le n.\end{aligned}\tag{2.6}$$

Note that the recurrence relation in (2.6) was first appeared in Ragab and Green (1984).

II. If $r=1$ in (2.3), then

$$\begin{aligned}\alpha_{1:n}^{(k)} &= \frac{-\pi\,\csc\frac{k\pi}{\alpha}}{(n-1)(n-2)!}\left(n-1-\frac{k}{\alpha}\right)\prod_{i=1}^{n-1}\left(i-1-\frac{k}{\alpha}\right)\\ &= \left[1-\frac{k}{\alpha(n-1)}\right]\alpha_{1:n-1}^{(k)},\quad n\ge 2.\end{aligned}\tag{2.7}$$

The recurrence relation in (2.7) first appeared in Ali and Khan (1987).

III. For $m\in\mathbb{N}$, (2.3) implies

$$\begin{aligned}\alpha_{r:n}^{(k-m\alpha)} &= \frac{(-1)^{r}\pi\,\csc\left(\frac{k}{\alpha}-m\right)\pi}{(r-1)!(n-r)!}\prod_{i=1}^{n}\left(i-r-\frac{k}{\alpha}+m\right)\\ &= \frac{(r-m-1)!(n-r+m)!}{(r-1)!(n-r)!}\,\frac{(-1)^{r-m}\pi\,\csc\frac{k\pi}{\alpha}}{(r-m-1)!(n-r+m)!}\\ &\quad\times\prod_{i=1}^{n}\left(i-(r-m)-\frac{k}{\alpha}\right)\\ &= \frac{(r-m-1)!(n-r+m)!}{(r-1)!(n-r)!}\,\alpha_{r-m:n}^{(k)},\quad m+1\le r\le n.\end{aligned}\tag{2.8}$$

When $m=1$, (2.8) reduces to the recurrence relation given by Ali and Khan (1987) as $\alpha_{r:n}^{(k-\alpha)}=\frac{n-r+1}{r-1}\,\alpha_{r-1:n}^{(k)},\quad 2\le r\le n.$

IV. Another form of (2.8) can be derived as follows

$$\begin{aligned}\alpha_{r:n}^{(k-m\alpha)} &= \frac{(-1)^{r+m}\pi\,\csc\frac{k\pi}{\alpha}}{(r-1)!(n-r)!}\prod_{i=1}^{n}\left(i+m-r-\frac{k}{\alpha}\right)\\ &= \frac{(-1)^{r+m}\pi\,\csc\frac{k\pi}{\alpha}}{(r-1)!(n-r)!}\prod_{i=m+1}^{n+m}\left(i-r-\frac{k}{\alpha}\right)\\ &= \frac{(-1)^{m}\prod_{i=n+1}^{n+m}\left(i-r-\frac{k}{\alpha}\right)}{\prod_{i=1}^{m}\left(i-r-\frac{k}{\alpha}\right)}\,\alpha_{r:n}^{(k)},\quad m+1\le r\le n.\end{aligned}\tag{2.9}$$

V. From (2.8) and (2.9), we get

$$\alpha_{r:n}^{(k)} = \frac{(-1)^m\,(r-m-1)!(n-r+m)!\,\prod_{i=1}^{m}\left(i-r-\frac{k}{\alpha}\right)}{(r-1)!(n-r)!\,\prod_{i=n+1}^{n+m}\left(i-r-\frac{k}{\alpha}\right)}\alpha_{r-m:n}^{(k)}, \tag{2.10}$$

$$+1 \le r \le n.$$

3. COVARIANCE BETWEEN ORDER STATISTICS

To calculate the covariance between $X_{r:n}$ and $X_{s:n}$, consider the joint pdf of $X_{r:n}$ and $X_{s:n}$, $1 \le r < s \le n$ as follows

$$f_{r,s:n}(x,y) = \alpha^2\, C_{r,s:n}\, \frac{x^{\alpha r-1}y^{\alpha-1}(y^\alpha - x^\alpha)^{s-r-1}}{(1+x^\alpha)^s(1+y^\alpha)^{n-r+1}}, \quad 0 \le x \le y < \infty, \tag{3.1}$$

where $C_{r,s:n} = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$.
Therefore the product moments, $\alpha_{r,s:n} = E(X_{r:n}Y_{s:n})$, can be written as

$$\alpha_{r,s:n} = \alpha^2\, C_{r,s:n} \int_0^\infty \int_0^y \frac{x^{\alpha r}y^\alpha(y^\alpha - x^\alpha)^{s-r-1}}{(1+x^\alpha)^s(1+y^\alpha)^{n-r+1}}dxdy. \tag{3.2}$$

On using the substitution $u = x^\alpha$ and $v = y^\alpha$, (3.2) reduces to

$$\alpha_{r,s:n} = C_{r,s:n} \int_0^\infty \frac{v^{\frac{1}{\alpha}}}{(1+v)^{n-r+1}} \underbrace{\left(\int_0^v \frac{u^{r+\frac{1}{\alpha}-1}(v-u)^{s-r-1}}{(1+u)^s}du\right)}_{I} dv. \tag{3.3}$$

By using the substitution $t = \frac{u}{v}$, it is not difficult to show that I can be simplified to

$$I = v^{s+\frac{1}{\alpha}-1} B\left(r+\frac{1}{\alpha}, s-r\right) {}_2F_1\left(s, r+\frac{1}{\alpha}, s+\frac{1}{\alpha}; -v\right), \tag{3.4}$$

where ${}_pF_q$ is the generalized hypergeometric function defined as

$${}_pF_q(a_1,...,a_p;b_1,...,b_q;x) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k}\,\frac{x^k}{n!}.$$

Using the Pfaff transformation, ${}_2F_1(a,b;c;x) = (1-x)^{-a}\,{}_2F_1(a,c-b;c;\frac{x}{x-1})$, we have

$${}_2F_1\left(s, r+\frac{1}{\alpha}, s+\frac{1}{\alpha}; -v\right) = (1+v)^{-r-\frac{1}{\alpha}}\,{}_2F_1\left(\frac{1}{\alpha}, r+\frac{1}{\alpha}, s+\frac{1}{\alpha}; \frac{v}{1+v}\right). \tag{3.5}$$

Now, using (3.4), (3.5) and the substitution $w = \frac{v}{1+v}$, (3.3) reduces to

$$\alpha_{r,s:n} = C_{r,s:n}\, B\left(r+\frac{1}{\alpha}, s-r\right) \int_0^1 w^{s+\frac{2}{\alpha}-1}(1-w)^{n-s-\frac{1}{\alpha}}\,{}_2F_1\left(\frac{1}{\alpha}, r+\frac{1}{\alpha}, s+\frac{1}{\alpha}; w\right). \tag{3.6}$$

On using the identity [Gradshteyn and Ryzhik, [14], p. 813]

$$\int_0^1 x^{\rho-1}(1-x)^{\sigma-1}\,{}_2F_1(\alpha,\beta,\gamma;x)dx = B(\rho,\sigma){}_3F_2(\alpha,\beta,\rho;\gamma,\rho+\sigma;1),$$

the product moments of the log-logistic distribution can be written as

$$\begin{aligned}\alpha_{r,s:n} &= C_{r,s:n}\,B\left(r+\frac{1}{\alpha},s-r\right)\;B\left(s+\frac{2}{\alpha},n-s-\frac{1}{\alpha}+1\right)\\ &\quad\times{}_3F_2\left(\frac{1}{\alpha},r+\frac{1}{\alpha},s+\frac{2}{\alpha};s+\frac{1}{\alpha},n+\frac{1}{\alpha}+1;1\right).\end{aligned}\tag{3.7}$$

It is clear from (3.7) that $\alpha_{r,s:n}$ exists for all $\alpha > 1$.

It is noteworthy to mention that one can use some existing recurrence relations in the literature to compute $\alpha_{r,s:n}$ in a more efficient way. For example, Joshi and Balakrishnan (1982) show that for any continuous distribution, the following recurrence relation holds

$$\begin{aligned}\alpha_{r,n:n} &= \sum_{i=1}^{n-r}(-1)^{n-r-i}\binom{n}{n-i}\binom{n-i-1}{r-1}\alpha_{n-i:n-i}\alpha_{i:i}\\ &\quad-\sum_{\ell=0}^{r-1}(-1)^{n-\ell}\binom{n}{\ell}\alpha_{1,n-r+1:n-\ell},\quad 1\le r\le n-1.\end{aligned}\tag{3.8}$$

Also, Ali and Khan (1987) show the following recurrence relation for the log-logistic distribution,

$$\begin{aligned}\alpha_{r,s:n} &= \alpha_{r,s-1:n}+\left(\frac{n}{n-s+1}\right)\left(1-\frac{1}{\alpha(n-s)}\right)\alpha_{r,s:n-1}\\ &\quad-\frac{n}{n-s+1}\alpha_{r,s-1:n-1},\quad 1\le r<s\le n-1.\end{aligned}\tag{3.9}$$

The covariance $\beta_{r,s:n} = \alpha_{r,s:n} - \alpha_{r:n}\,\alpha_{s:n}$, can be obtained from equations (2.4), (2.5) and (3.7). Note that when $r = s$, the variances $\beta_{r,r:n} = \alpha^2_{r:n} - (\alpha_{r:n})^2$. The recurrence relations in (23) and (24) can be also used in these calculations.

4. PARAMETER ESTIMATION FOR THE LOG-LOGISTIC DISTRIBUTION

In this section, we discuss the parameter estimation for the log-logistic distribution based on order statistics.

4.1. Estimation of location and scale parameters

Let $Y_1, Y_2, ..., Y_n$ be a random sample of size n from the log-logistic$(\alpha, \theta_1, \theta_2)$, where θ_1 is the location parameter and $\theta_2 > 0$ is the scale parameter. I.e. $f(y) = \alpha\,\theta_2^{-1}\left(\frac{y-\theta_1}{\theta_2}\right)^{\alpha-1}\left(1+\left(\frac{y-\theta_1}{\theta_2}\right)^{\alpha}\right)^{-2}$, $y \geq \theta_1$. In this section, we compute the best linear unbiased estimators (BLUEs) for θ_1 and θ_2 when the shape parameter α is known. Let $X = (Y-\theta_1)/\theta_2$. When α is known, the mean, $\alpha_{r:n}$, and the covariance, $\beta_{r,s:n}$, of order statistics are completely known and free of parameters. The estimators for θ_1 and θ_2 are derived based on weighted regression on the quantile-quantile plot of order statistics against their expected value. The weights depend on the covariance matrix of the order statistics. The estimations of location and scale parameters based on order statistics were originally introduced by Lloyd [20]. Several authors including Arnold *et al.* ([6], p. 17) and Ahsanullah *et al.* ([3], p. 154) used Lloyd's method to obtain best linear unbiased estimator (BLUE) of the location and scale parameters for probability distributions. The BLUEs of θ_1 and θ_2 can be computed as follows [see Arnold *et al.* ([6], pp. 171–173) and Ahsanullah *et al.* ([3], p. 154)]

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{A'}\,\boldsymbol{\Sigma^{-1}}\,\boldsymbol{A})^{-1}\,\boldsymbol{A'}\,\boldsymbol{\Sigma^{-1}}\,\boldsymbol{Y},$$

where $\boldsymbol{A'}$ denotes to the transpose of $\boldsymbol{A}$, $\boldsymbol{A} = (\mathbf{1}, \boldsymbol{\mu})$, $\mathbf{1} = (1, 1, ..., 1)'_{1\times n}$, $\boldsymbol{\mu} = (\alpha_{1:n}, \alpha_{2:n}, ..., \alpha_{n:n})'$, $\hat{\boldsymbol{\theta}} = (\hat{\theta_1}, \hat{\theta_2})'$, $\boldsymbol{\Sigma} = ((\beta_{r,s:n}))_{n\times n}$ and $\boldsymbol{Y} = (Y_{1:n}, Y_{2:n}, ..., Y_{n:n})'$. Alternatively,

$$\hat{\theta_1} = -\boldsymbol{\mu}'\,\boldsymbol{\Gamma}\,\boldsymbol{Y} \quad \text{and} \quad \hat{\theta_2} = \mathbf{1}'\,\boldsymbol{\Gamma}\,\boldsymbol{Y},$$

where $\boldsymbol{\Gamma} = \boldsymbol{\Sigma^{-1}}(\mathbf{1}\boldsymbol{\mu}' - \boldsymbol{\mu}\,\mathbf{1}')\boldsymbol{\Sigma^{-1}}/\boldsymbol{\Delta}$ and $\boldsymbol{\Delta} = (\mathbf{1}'\boldsymbol{\Sigma^{-1}}\mathbf{1})(\boldsymbol{\mu}'\boldsymbol{\Sigma^{-1}}\boldsymbol{\mu}) - (\mathbf{1}'\boldsymbol{\Sigma^{-1}}\boldsymbol{\mu})^2$. The coefficient matrix $\boldsymbol{C} = (\boldsymbol{A'}\,\boldsymbol{\Sigma^{-1}}\,\boldsymbol{A})^{-1}\,\boldsymbol{A'}\,\boldsymbol{\Sigma^{-1}}$ can be obtained using $\alpha_{r:n}$, $\alpha^{(2)}_{r:n}$, $\alpha_{r,s:n}$ and $\beta_{r,s:n}$ from previous section. The covariance matrix of the estimators can be computed in terms of θ_2 as follows

$$\boldsymbol{Cov}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{A'}\,\boldsymbol{\Sigma^{-1}}\,\boldsymbol{A})^{-1}\theta_2^2. \tag{4.1}$$

In particular,

$$\begin{aligned} var(\hat{\theta_1}) &= \theta_2^2\boldsymbol{\mu'\Sigma^{-1}}\,\boldsymbol{\mu}/\boldsymbol{\Delta}, \\ var(\hat{\theta_2}) &= \theta_2^2\mathbf{1'\Sigma^{-1}}\,\mathbf{1}/\boldsymbol{\Delta}, \\ Cov(\hat{\theta_1}, \hat{\theta_2}) &= -\theta_2^2\boldsymbol{\mu'\Sigma^{-1}}\mathbf{1}/\boldsymbol{\Delta}. \end{aligned}$$

Equation (4.1) is used to compute the variance and covariance of $\hat{\theta_1}$ and $\hat{\theta_2}$ in terms of θ_2. The coefficients and covariances for computing the BLUE of $\hat{\boldsymbol{\theta}}$ for various values of the shape parameter α and sample sizes up to 10 are available on `https://sites.google.com/site/statisticsmanagementservices/`.

4.2. Estimation of the shape parameter

In real life situations, we encounter unknown value for the shape parameter α. In order to use the The BLUEs for θ_1 and θ_2, we first estimate the shape parameter α.

Lemma 4.1. *The log-logistic distribution is a member of the Pareto-type distributions with tail index α.*

Proof: Note that $1 - F(x) = \frac{1}{1+x^{\alpha}} = x^{-\alpha}\ell(x)$, where $\ell(x) = 1 - x^{-\alpha} + x^{-2\alpha} + \cdots$ is slowly varying function at infinity. To see this, for any $\lambda > 0$, $\frac{\ell(\lambda\, x)}{\ell(x)} \longrightarrow 1$ as $x \longrightarrow \infty$. Hence $F(x)$ constitutes a Pareto-type distribution with tail index α. □

Several estimators for the heavy tail index α exist in the literature. For example, a family of kernel estimators for α was proposed by Csorgo, Deheuvels and Mason [12]. Bacro and Brito [8] and De Hann [13] proposed estimators for α which are members of the family of kernel estimators. For more information, we refer the reader to the paper by Beirlant *et al.* [9] and Gomes and Henriques-Rodrigues [17]. The most popular estimator for α is the Hill estimator proposed by Hill [18] as follows:
Let $X_1, X_2, ..., X_n$ be n independent random sample from log-logistic$(\alpha, \theta_1, \theta_2)$. Let $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ be the corresponding order statistics. The Hill estimator for α based on upper k order statistics is given by

$$\hat{\alpha} = \frac{1}{H_{k,n}}, \quad H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log \frac{X_{n-j+1,n}}{X_{n-k,n}}. \tag{4.2}$$

Although the Hill estimator is scale invariant, it is not shift invariant. Aban and Meerschaert [1] proposed a modification of Hill estimator in order to make it both shift and scale invariant as follows:

$$\hat{\alpha}^{-1} = \frac{1}{k}\sum_{j=1}^{k} \log \frac{X_{n-j+1,n} - \hat{s}}{X_{n-k,n} - \hat{s}},$$

where the sift $\hat{s}$ satisfies the equation

$$\hat{\alpha}(X_{n-k,n} - \hat{s})^{-1} = \frac{\hat{\alpha} + 1}{k}\sum_{j=1}^{k}(X_{n-j+1,n} - \hat{s})^{-1}, \quad \hat{s} < X_{n-k,n}.$$

In general, the modified Hill estimator results in large variation of the sampling distribution in compared with the Hill estimator. In our case, based

on various simulated random samples with different sample sizes from $X \sim$ log-logistic$(\alpha, \theta_1, \theta_2)$, the modified Hill estimator produces poor estimate for the parameter α. Therefore, we decided to shift the random sample by the sample minimum, $X_{1,n}$, and then use the Hill estimator to estimate α. This is justified since the lower end of the distribution is finite. For an interesting discussion of this topic see Araújo-Santos *et al.* [5] and Gomes *et al.* [15]. The results showed good estimate to the shape parameter α (see Table 1).

4.3. Monte Carlo simulation study

In this subsection, we generate different random samples with various sizes, $n = 100,\ 500,\ 1,000$ and $10,000$. The simulation study is repeated 1,000 times for four groups of parameters:

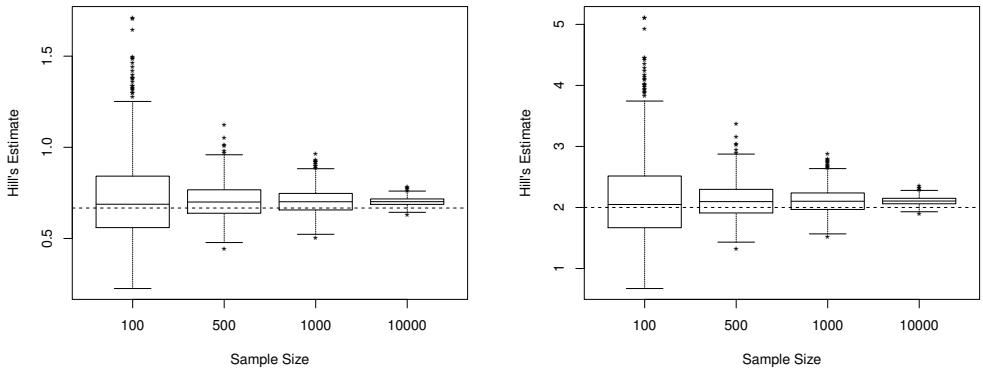
$$\begin{aligned} \text{I}: &\ \alpha = 0.5,\ \theta_1 = 1,\ \theta_2 = 1, \\ \text{II}: &\ \alpha = 1.5,\ \theta_1 = 0,\ \theta_2 = 1, \\ \text{III}: &\ \alpha = 2.5,\ \theta_1 = 2,\ \theta_2 = 3, \\ \text{IV}: &\ \alpha = 4, \quad \theta_1 = 2,\ \theta_2 = 0.5. \end{aligned}$$

For each parameter combination, we generate random samples, Y_i, $i = 1, ..., n$ from log-logistic$(\alpha, \theta_1, \theta_2)$. We assume the random sample $X_i = Y_i - Y_{1,n}$ follows log-logistic$(\alpha, 0, \theta_2)$. Then we estimate α using the Hill estimator in equation (4.2). Gomes and Guillou [16] have given an interesting discussion about the choice of k. It is known that the bias of the estimator of the index parameter increases as k increases and the variance of the index estimator increases if k is small. The choice of k is a question between the choice of bias and variance.

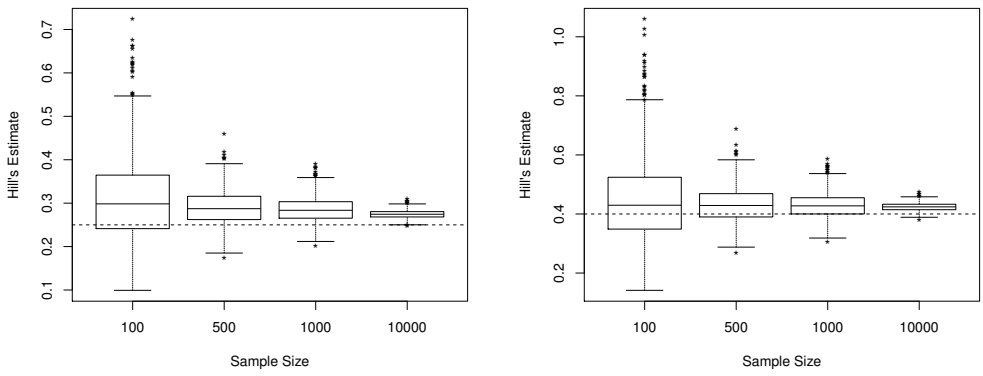
Table 1: Mean, median and standard deviation for $\hat{\alpha}$ using the Hill estimate.

Sample Size	Summary Statistics	Group I	Group II	Group III	Group IV
		$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\alpha}$	$\hat{\alpha}$
100	Median	0.4881	1.4539	2.3275	3.3510
	Mean	0.5158	1.5371	2.4615	3.5460
	Standard Deviation	0.1716	0.5094	0.8067	1.1494
500	Median	0.4771	1.4287	2.3320	3.4810
	Mean	0.4827	1.4450	2.3600	3.5150
	Standard Deviation	0.0690	0.2063	0.3350	0.4995
1000	Median	0.4758	1.4256	2.3390	3.5260
	Mean	0.4786	1.4342	2.3540	3.5460
	Standard Deviation	0.0473	0.1418	0.2330	0.3584
10000	Median	0.4750	1.4246	2.3580	3.6438
	Mean	0.4754	1.4258	2.3616	3.6460
	Standard Deviation	0.0147	0.0441	0.0733	0.1211

We have taken $k = 10\%$ of the sample size with $n > 100$. The simulation results in Table 1 show that as the parameter α increases, the absolute bias and standard deviation increase. Overall, the Hill estimator performs well in estimating the shape parameter α. Figures 1–4 represent the Boxplots for the observed sampling distributions of the Hill estimate for different sample sizes. These Figures indicate that the observed distributions are approximately normal and centered roughly at α^{-1}.



(a) Group I (b) Group II



(a) Group III (b) Group IV

Figure 1: Boxplots for the observed sampling distributions of $\hat{\alpha}^{-1}$. Dashed line represents the true parameter α^{-1}.

4.4. Numerical Example

In this subsection, we illustrate the use of Hill estimator and the BLUE's for estimating the three-parameter log-logistic distribution. We simulate a random sample with $n = 30$ observations from log-logistic distribution with parameters $\alpha = 4$, $\theta_1 = 2$ and $\theta_2 = 3$. The simulated data is given below:

5.80310, 6.88820, 6.00730, 7.01140, 4.87250, 4.00560, 4.49970, 5.02880, 5.83690, 11.40110, 3.30511, 3.95312, 5.87513, 2.55114, 4.68615, 4.88916, 4.67717, 4.71818, 4.05190, 8.31920, 4.86421, 4.50422, 8.89623, 5.74124, 5.48125, 4.68226, 5.70127, 5.13528, 4.20729, 4.95430.

Using similar approach as in subsection 4.3, the estimated value of α based on the Hill estimator is $\hat{\alpha} = 3.350$. Based on this value and the sample size of $n = 30$, the coefficient matrix, $\boldsymbol{C} = (\boldsymbol{A'\,\Sigma^{-1}\,A})^{-1}\,\boldsymbol{A'\,\Sigma^{-1}}$, and the covariance, $\boldsymbol{Cov}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{A'\,\Sigma^{-1}\,A})^{-1}\theta_2^2$, can be calculated using $\alpha_{r:n}$, $\alpha_{r:n}^{(2)}$, $\alpha_{r,s:n}$ in equations (9), (10) and (22) respectively. These coefficients for computing the BLUE's for θ_1 and θ_2 and the covariance matrix are provided below:

$$\boldsymbol{C_{\theta_1}} = \begin{pmatrix} 0.6077 \\ 0.25102 \\ -0.08301 \\ 0.66281 \\ -0.17817 \\ -0.09813 \\ 0.27104 \\ -0.02089 \\ 0.0082 \\ 0.07347 \\ -0.18869 \\ 0.48597 \\ -0.09435 \\ -0.48662 \\ -0.0126 \\ 0.51641 \\ -0.302 \\ 0.43571 \\ -0.66808 \\ 0.0207 \\ 0.11805 \\ -0.16527 \\ 0.00062 \\ -0.01915 \\ -0.07406 \\ 0.00442 \\ -0.00919 \\ -0.04542 \\ -0.00744 \\ -0.00303 \end{pmatrix}, \qquad \boldsymbol{C_{\theta_2}} = \begin{pmatrix} -0.63961 \\ -0.23399 \\ 0.09682 \\ -0.64294 \\ 0.15847 \\ -0.01285 \\ 0.12400 \\ 0.00067 \\ -0.00583 \\ -0.01679 \\ 0.15861 \\ -0.09637 \\ 0.00875 \\ 0.44036 \\ 0.02531 \\ -0.46079 \\ 0.23208 \\ -0.21193 \\ 0.70721 \\ -0.02347 \\ -0.04269 \\ 0.18111 \\ -0.01123 \\ 0.07461 \\ 0.08937 \\ -0.00810 \\ 0.05792 \\ 0.03363 \\ 0.01434 \\ 0.00333 \end{pmatrix},$$

$$\boldsymbol{Cov}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} 0.01636 & -0.01755 \\ -0.01755 & 0.02777 \end{pmatrix}.$$

Therefore, the BLUE's for θ_1 and θ_2 and the estimated covariances are evaluated to be

$$\hat{\theta_1} = 1.98287, \quad \hat{\theta_2} = 3.09528,$$

$$Var(\hat{\theta_1}) = 0.15676, \quad Var(\hat{\theta_2}) = 0.26606$$

and

$$Cov(\hat{\theta_1}, \hat{\theta_2}) = -0.16812.$$

5. CONCLUDING REMARKS

In this paper, the moments and product moments of the order statistics in a sample of size n drawn from the log-logistic distribution are discussed. We also provided in the same section more compact formulas for the means, variances and covariances of order statistics. Best linear unbiased estimators (BLUEs) for the location and scale parameters for the log-logistic distribution with known shape parameter based on order statistics are studied. The Hill estimator is proposed for estimating the shape parameter.

ACKNOWLEDGMENTS

The authors are grateful for the comments and suggestions by the referee and the Associate Editor. Their comments and suggestions greatly improved the article.

REFERENCES

[1] ABAN, I.B. and Meerschaert, M.M. (2001). Shifted Hill's estimator for heavy tails, *Communications in Statistics-Simulation and Computation*, **30**, 949–962.

[2] Abramowitz, M. and Stegun, I.A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, Vol. 55, Dover Publication, Inc., New York.

[3] Ahsanullah, M.; Nevzorov, V.B. and Shakil, M. (2013). *Order statistics*, Atlantis Press, Paris, France.

[4] Ali, M.M. and Khan, A.H. (1987). On order statistics from the log-logistic distribution, *Journal of Statistical Planning and Inference*, **17**, 103–108.

[5] Araújo-Santos, P.; Fraga Alves, M.I. and Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and high quantile estimation, *Revstat*, **4**, 227–247.

[6] Arnold, B.C.; Balakrishnan, N. and Nagaraja, H.N. (1992). *A first course in order statistics*, Vol. 54, SIAM.

[7] Ashkar, F. and Mahdi, S. (2006). Fitting the log-logistic distribution by generalized moments, *Journal of Hydrology*, **238**, 694–703.

[8] Bacro, J.N. and Brito, M. (1993). Strong limiting behaviour of a simple tail Pareto-index estimator, *Statistics and Decisions*, **3**, 133–134.

[9] Beirlant, J.; Vynckier, P. and Teugels, J.L. (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics, *Journal of the American statistical Association*, **91**, 1659–1667.

[10] Bennett, S. (1983). Log-Logistic Regression Models for Survival Data, *Journal of the Royal Statistical Society, Series C*, **32**, 165–171.

[11] Collett, D. (2003). *Modelling Survival Data in Medical Research*, 2nd edition, CRC Press.

[12] Csorgo, S.; Deheuvels, P. and Mason, D. (1985). Kernel estimates of the tail index of a distribution, *The Annals of Statistics*, **13**, 1050–1077.

[13] De Haan, L. (1981). Estimation of the minimum of a function using order statistics, *Journal of the American Statistical Association*, **76**, 467–469.

[14] Gradshteyn, I.S. and Ryzhik, I.M. (2007). *Table of Integrals, Series and Products*, 7th edition, Academic, New York, USA.

[15] Gomes, M.I.; Fraga Alves, M.I. and Araújo-Santos, P. (2008). PORT Hill and moment estimators for heavy-tailed models, *Communications in Statistics-Simulation and Computation*, **37**, 1281–1306.

[16] Gomes, M.I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: A review, *International Statistical Review*, **83**, 263–292.

[17] Gomes, M.I. and Henriques-Rodrigues, L. (2016). Competitive estimators of extreme value Index, *Statistics and Probability Letters*, **67**, 128–135.

[18] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, **3**, 1163–1174.

[19] Joshi, P.C. and Balakrishnan, N. (1982). Recurrence relations and identities for the product moments of order statistics, *Sankhya: The Indian Journal of Statistics, Series B*, **44**, 39–49.

[20] Lloyd, E.H. (1952). Least squares estimation of location and scale parameters using order statistics, *Biometrika*, **39**, 88–95.

[21] Ragab, A. and Green J. (1984). On order statistics from the log-logistic distribution and their properties, *Communications in Statistics-Theory and Methods*, **13**, 2713–2724.

[22] Shah, B.K. and Dave, P.H. (1963). A note on log-logistic distribution, *Journal of the MS University of Baroda*, **12**, 15–20.

[23] Shoukri, M.M.; Mian, I.U.M. and Tracy, D.S. (1988). Sampling Properties of Estimators of the Log-Logistic Distribution with Application to Canadian Precipitation Data, *The Canadian Journal of Statistics*, **16**, 223–236.

[24] Tadikamalla, P.R. and Johnson, N.L. (1982). Systems of frequency curves generated by transformations of logistic variables, *Biometrika*, **69**, 461–465.

[25] Voorn, W.J. (1987). Characterization of the logistic and log-logistic distributions by extreme value related stability with random sample size, *Journal of Applied Probability*, **24**, 838–851.

REVSTAT – Statistical Journal
Volume 16, Number 4, October 2018, 445–462

BAYESIAN ROBUSTNESS MODELLING USING THE FLOOR DISTRIBUTION

Authors: J.A.A. Andrade
– Department of Statistics and Applied Mathematics,
Federal University of Ceara, 60455-670, Brazil
ailton@ufc.br

Edward Omey
– Mathematics and Statistics,
HUB, Stormstraat 2, 1000 Brussels, Belgium
edward.Omey@kuleuven.be

C.T.M. Aquino
– Department of Statistics and Applied Mathematics,
Federal University of Ceara, 60455-670, Brazil
mendes.aquino@gmail.com

Received: January 2016 Revised: October 2016 Accepted: October 2016

Abstract:

- In Bayesian analysis, the prior distribution and the likelihood can conflict, in the sense that they can carry diverse information about the parameter of interest. The most common form of conflict is the presence of outliers in the data. Usually, problems of conflicts are solved by assigning heavy-tailed distributions to that source of information which may be causing the conflict. However, the class of heavy-tailed distributions is not well defined, therefore there are many ways to define heavy tails. The class O-regularly varying distributions is rather unknown in Statistics, it basically embraces those distributions whose tails decay oscillating between two power functions. In this work we study a new distribution which has this property and, as a consequence, yields robust models for location and for scale parameter models separately. We provide explicit expressions for some relevant quantities concerning the distribution, such as the moments, distribution function, etc. Besides, we show how conflicts can be resolved using this distribution.

Key-Words:

- *Bayesian robustness modelling; conflicting information; O-regularly varying distributions; heavy-tailed distributions.*

AMS Subject Classification:

- 62E15.

1. INTRODUCTION

In Bayesian analysis, two sources of information are used to study a phenomenon of interest: the prior distribution and the likelihood function. In this way we combine, through Bayes' theorem, the evidences from both the data and some relevant subjective knowledge about the parameter of interest. However, since models are only a try to describe the reality, which is much more complex, the modelling process is inevitably subject to errors. In fact, the model misspecification can lead to wrong conclusions, since it can strongly affect the posterior distribution. The conflict of information can arise from a model not prepared to deal with diverse information, e.g. outliers, leading to different evidences about the parameters, one coming from the prior and the other coming from the likelihood. The most common form of conflict are the outliers, since they will carry information far apart from the prior distribution and the rest of the data.

This behaviour was first identified by Lindley (1968), who suggest the use of the Student-t distribution to resolve the conflicts. Dawid (1973) established conditions on the data and the prior distributions which yields robust posterior distributions for the location parameters. Several works followed this thread, basically improving Dawid's conditions as, for instance, O'Hagan (1979, 1988 and 1990) and O'Hagan & Le (1994). All these works concerned only location parameter models. In order to solve conflict of information in scale and location parameter structures separately, Andrade & O'Hagan (2006) study a class of heavy-tailed distributions different from the ones considered by Dawid (1973) and O'Hagan (1979), namely the class of *regularly varying distributions*. A more general approach for location-scale structures was proposed by Andrade & O'Hagan (2011). Andrade *et al.* (2013) proposed alternative conditions for the location and the scale parameter models which are slightly easier to verify. Andrade & Omey (2013) give several new conditions using different classes of distributions, such as the subexponential and L classes. For a more complete literature review on robustness modelling, see O'Hagan & Pericchi (2012). The papers of Andrade & O'Hagan showed that, working within the regularly varying class, the outlying information will be only partially rejected, in the sense that it exerts an initial influence which, even though is constant, does not vanish as the outlier becomes large. Thus, concerning this aspect Desgagné (2013, 2015) proposes a new class of distributions which allow to resolve conflicts in scale (and location-scale) parameter(s) models by fully rejecting the conflicting information (full robustness). However, this led to rather complex conditions and distributions, which can limit the applications. Andrade & Omey (2016) proposes to use the class of *O-regularly varying distributions* (ORV), which are much more intuitive and also allows full robustness in scale parameter models.

The robustness we are treating is related to the conflict of information, the thread initiated by Dawid (1973), in which some of the sources of information

(prior/likelihood) carries some information that is away from the rest of the information (See Andrade & O'Hagan, 2006).

One important thing to notice is that, although there are many new distributions which could be used for robustness modelling, for the best of our knowledge, no other distribution has this tail behaviour, with oscillating decay between two regularly varying distributions. This peculiar behaviour of the floor distribution, besides being heavy tail, also can motivate the creation of new distributions involving some sort of waving functions.

In this work we study the Floor distribution, which is in the O-regularly varying class of heavy-tailed distributions. It was firstly suggested by Andrade & Omey (2016), however here we compute all the relevant quantities of the distribution, such as moments, distribution function, random numbers generator, etc. In Section 2 we give the definitions of regular and O-regular variation. In addition we provide a rule for creating ORV distributions and show how conflicts of information can be resolved in Bayesian analysis context. In Section 3 we present the Floor distribution and its quantities. A simulation study comparing the floor distribution with the exponential one, using different proportions of outliers and sample sizes, is provided in Section 4. Section 5 provides an example in which outlying information is automatically rejected by the model as it becomes large. Finally, we conclude with some general remarks in Section 6.

2. HEAVY TAILS AND REGULAR VARIATION

Roughly speaking the concept of heavy tails is associated with those distributions whose tails decay at least slower than the function e^{-x}. However, there is not a widely spread accepted definition of heavy tail. Most of the definitions in the literature are contextualised in some area. In this work we define heavy tail as regular variation.

The concept of regular variation was introduced by Karamata (1930). Feller (1971) studied the application of such concept on probability theory and Andrade & O'Hagan (2006 e 2011) used regularly varying distributions in robust Bayesian modeling. Others studies include Landau (1911), Valiron (1913), Pólya (1917), de Haan (1970) and Seneta (1976). The main reference about regular variation used in this paper is the book written by Bingham *et al.* (1987).

Definition 2.1. (Regular variation) A measurable function f is said to be regularly varying at infinity with index ρ, $\rho \in \mathbb{R}$ and $\lambda > 0$, if

$$\lim_{x\to\infty} \frac{f(\lambda\, x)}{f(x)} = \lambda^{\rho}. \tag{2.1}$$

We denote it by $f \in R_{\rho}$.

Particularly, if $\rho = 0$, then f is said to be a slowly varying function, denoted by $f \in R_0$. We write the set of all regularly varying functions as $R = \{R_\rho : \rho \in (-\infty, \infty)\}$. The characterisation theorem establishes that if $f(x) \in R_\rho$, then $f(x)$ can be written as $f(x) = x^\rho \ell(x)$, where $\ell(x)$ is a slowly varying function. For more details about Karamata's theory and, in particular, regular variation, see Bingham *et al.* (1987). Definition 2.1 can be interpreted as the tail behavior of a probability density function, i.e., there is a relation between regular variation and heavy tails. For example, if $f \in R_{-\rho}$ as $x \to \infty$, then the right tail decreases like a power function $x^{-\rho}$.

The O-regular variation class (which we call ORV) extends the concept of regular variation and was introduced by Avakumović (1936). This class involves distributions whose tails decrease at any behaviour between two regularly varying functions. For example, the tails of an ORV distribution can decrease oscillating between two power functions (Andrade & Omey, 2016).

Definition 2.2 (O-regular variation)**.** A probability density f is said to be O-regularly varying at *infinity*, denoted by $f \in ORV$, if f satisfies

$$\lim_{x\to\infty} \sup \frac{f(xy)}{f(x)} < \infty, \forall y > 0. \tag{2.2}$$

Along with the definition of ORV, we have the *upper* and *lower Matuszewska indexes*. If $f \in ORV$, the upper index of f is given by

$$\alpha(f) = \lim_{y\to\infty} \frac{\log \limsup_{x\to\infty} f(xy)/f(x)}{\log(y)}, \tag{2.3}$$

and the lower index of f is given by

$$\beta(f) = \alpha(1/f) = \lim_{y\to\infty} \frac{\log \liminf_{x\to\infty} f(xy)/f(x)}{\log(y)}. \tag{2.4}$$

Since the tails of ORV distributions decrease between two polynomials, it is a broad class, thus there are many ways of constructing ORV distributions. Andrade & Omey (2016) suggest a procedure to create ORV distributions. Let f be a probability density function of the form

$$f(x) = Cb(x)A(x), \tag{2.5}$$

where C is the normalizing constant, b is bounded away from zero to infinity when x tends to infinity. For the class of distributions defined in (5), it follows that:

- (**i**) If $xA'(x)/A(x)$ is bounded, then $f \in$ ORV;
- (**ii**) If $A(x) \in RV_{-\alpha}$, then $f \in$ ORV.

Andrade & Omey (2016) showed that the floor distribution belongs to the O-regular class. We say that the random variable X is distributed according to a floor distribution with parameter a, denoted by $X \overset{d}{\sim} \text{floor}(a)$, if its probability density function is given by

$$f(x) = C(a)\, x^{-a}\, e^{\lfloor \log x \rfloor}, 1 \leq x < \infty, \tag{2.6}$$

where $C(a)$ is the normalizing constant, $a > 2$ and $\lfloor . \rfloor$ is the floor function. In other words,

$$f(x) = C(a) x^{-a+1} e^{\lfloor \log x \rfloor - \log x},$$

for $x \geq 1$. Such authors also prove that the floor distribution belongs to the O-regularly varying class of distribution. Note that this satisfies (2.5), where $A(x) = x^{a-1}$ and $b(x) = e^{\lfloor \log x \rfloor - \log x}$. It is easy to see that $A(x)$ is regularly varying with index $-a+1$. Since $\lfloor \log x \rfloor \leq \log x \leq \lfloor \log x \rfloor + 1$, we also have that $-1 \leq \log x \leq 0$, which shows that $e^{-1} \leq b(x) \leq 1$. Thus we have that the floor distribution is a O-regularly varying. We also have that the Matuszewska indices for the floor distribution are given by $\alpha(f) = \beta(f) = -a + 1$.

Consider $\boldsymbol{x} = (x_1, ..., x_n)|\theta \overset{iid}{\sim} f(x|\theta) = \theta^{-1} h(x/\theta)$, $\theta \sim p(\theta)$, and h and p bounded continuous probability densities. Following the notation in Andrade & Omey (2016), the data are partitioned in two sets, called $\boldsymbol{x}^L$ and $\boldsymbol{x}^U$, defined by $L = f(\boldsymbol{x}^L|\theta) = \prod_{i=1}^{k} h(x_i|\theta)$ and $U = f(\boldsymbol{x}^U|\theta) = \prod_{i=k+1}^{n} h(x_i|\theta)$, where $\boldsymbol{x}^L$ are the outliers. In other words, $f(\boldsymbol{x}|\theta) = \theta^{-n} \times L \times U$, and the posterior distribution is given by

$$p(\theta|\boldsymbol{x}) = \frac{\theta^{-n} \times L \times U \times p(\theta)}{\int_0^\infty \theta^{-n} \times L \times U \times p(\theta) d\theta}.$$

Andrade & Omey (2016) showed that, if the following conditions hold,

- **(i)** $h \in ORV$ with $\alpha(h) < 0$,
- **(ii)** $\int_0^1 y^{-k\alpha - n} \times U \times p(y) dy < \infty$,
- **(iii)** $\int_1^\infty y^{-n-k\beta} \times U \times p(y) dy < \infty$,
- **(iv)** $\int_x^\infty y^{-n} p(y) dy = O(1) \Pi_{i=1}^{k} x_i^{\beta - \epsilon}$,

then

$$0 < \liminf_{x \to \infty} \frac{p(\theta|\boldsymbol{x})}{U\ p(\theta)} \leq \limsup_{x \to \infty} \frac{p(\theta|\boldsymbol{x})}{U\ p(\theta)} < \infty \tag{2.7}$$

The result in (2.7) establishes that, as x tends to infinity, the posterior distribution will be bounded by two quantities independent of x. Thus, the posterior distribution will be based on the prior information and the observations that are not outliers.

3. THE FLOOR DISTRIBUTION

To define the location-scale floor family of distributions, let us consider the linear transformation $Z = \sigma X + \mu$ where $\mu \in \mathbb{R}$, $\sigma > 0$ and $X \overset{d}{\sim} \text{floor}(a)$.

3.1. The probability density function

Consequently, the density of Z is given by

$$h(z) = C(a)\,\frac{1}{\sigma}\left(\frac{z-\mu}{\sigma}\right)^{-a} e^{\left\lfloor \log\left(\frac{z-\mu}{\sigma}\right)\right\rfloor}, \mu + \sigma \leq z < \infty,$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. Figure 1 shows plots of the density for different values of a, μ, and σ.

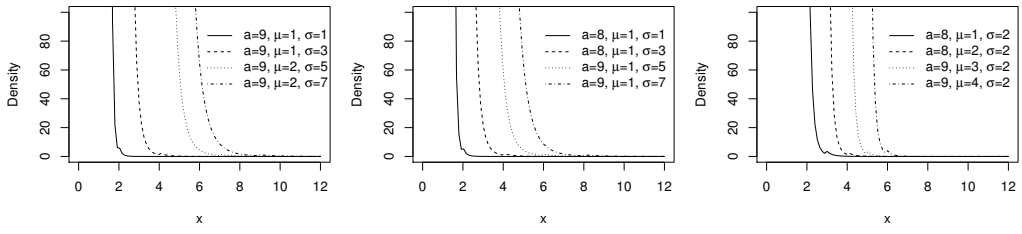


Figure 1: Floor density for different values of a, μ, e σ.

3.2. The normalising constant

Note that (2.6) will be a density function if

$$\int_1^\infty C(a)\, x^{-a}\, e^{\lfloor \log x \rfloor}\, \mathrm{d}x \;=\; 1. \tag{3.1}$$

Applying the transformation $t = \log(x)$,

$$\begin{aligned}
\frac{1}{C(a)} &= \int_0^\infty e^{-a\,t}\, e^{\lfloor t \rfloor}\, e^t \,\mathrm{d}t \\
&= \int_0^\infty e^{-t(a-1)}\, e^{\lfloor t \rfloor} \,\mathrm{d}t \\
&= \int_0^1 e^{-t(a-1)}\, e^0 \, dt + \int_1^2 e^{-t(a-1)}\, e^1 \,\mathrm{d}t + \cdots \qquad (3.2) \\
&= \sum_{n=0}^{\infty} e^n \int_n^{n+1} e^{-t(a-1)} \,\mathrm{d}t \\
&= \frac{1}{a-1}\left[\sum_{n=0}^{\infty} e^{-n(a-2)} - \sum_{n=0}^{\infty} e^{-(n+1)(a-1)+n}\right].
\end{aligned}$$

The two sums in (3.2) are geometric series with ratio $e^{-(a-2)}$, hence

$$C(a) = \frac{(a-1)(e^2 - e^a)}{e - e^a},\ a > 2.$$

3.3. Cumulative distribution function

The cumulative distribution function of $X \sim \text{Floor}(a)$ is given by:

$$F(x) \,=\, C(a) \int_1^x t^{-a}\, e^{\lfloor \log t \rfloor} \,\mathrm{d}t$$

Using the transformation $\nu = \log t$,

$$F(x) \,=\, \frac{C(a)}{a-1}\left[\sum_{n=0}^{\lfloor \log x \rfloor} e^{-n(a-2)} - e^{1-a} \sum_{n=0}^{\lfloor \log x \rfloor - 1} e^{-n(a-2)} - e^{(1-a)\log x + \lfloor \log x \rfloor}\right] \qquad (3.3)$$

The two sums in (3.3) are partial geometric series with ratios $e^{-(a-2)}$ e e^{-a}, respectively. It follows that

$$F(x) = -\frac{C(a)}{a-1}\left\{e^{-\log(x)(a-1)+\lfloor \log(x) \rfloor} + (e^{-1} - 1)\left[\frac{1 - e^{-(a-2)\lfloor \log x \rfloor}}{e^{a-2} - 1}\right] - 1\right\},$$

for $x > 1$ and $a > 2$. Note that the cumulative distribution involving the location and scale parameters can be obtained by simple variable transformation, that is letting $Z = \sigma X + \mu$, we have the expression $F_Z(z) = P(X \leq \frac{z-\mu}{\sigma})$.

3.4. Moment of order r

The moment of order r is calculated replacing a for $(a-r)$ in (3.1). Thus,

$$\mathbb{E}(X^r) = C(a)\int_1^\infty x^{-(a-r)}\, e^{\lfloor \log x \rfloor}\, \mathrm{d}x = \frac{C(a)}{C(a-r)}, \tag{3.4}$$

which exists only when $r < a-2$.

3.5. Summaries

Proposition 3.1. *If X is a random variable following the standard floor distribution with parameter a, then*

(**i**) *The expectation of X, given straightforwardly by (3.4), is given by* $\mathbb{E}(X) = C(a)/C(a-1),\ a \geq 3$.

(**ii**) *Variance. Using (3.4), the variance is*

$$Var(X) = \frac{C(a)[C^2(a-1) - C(a)C(a-2)]}{C(a-2)C^2(a-1)},\ a > 4. \tag{3.5}$$

(**iii**) *Quantiles can be obtained by numerically inverting the cumulative distribution function.*

(**iv**) *The coefficient of skewness is given by*

$$\begin{aligned}\gamma_1(X) \;=\; & \frac{C^3(a-1)C(a-2) - 3C(a)C^2(a-1)C(a-3) + 2C^2(a)C(a-2)C(a-3)}{C^{1/2}(a)C(a-3)[C^2(a-1) - C(a)C(a-2)]^{3/2}} \\ & \times C^{1/2}(a-2), \quad a > 5.\end{aligned}$$

(**v**) *It can be shown that the excess kurtosis can be written as*

$$\begin{aligned}K(X) \;=\; & C(a-1)C(a-2)\{C^3(a-1)\prod_{i=2}^{3} C(a-i) - 4C(a-4) \\ & \times \prod_{i=0}^{2} C(a-i) + 6C^2(a)C(a-1)\prod_{i=3}^{4} C(a-i) - 4C(a) \\ & \times \prod_{i=0}^{4} C(a-i) + C^2(a)\prod_{i=2}^{4} C(a-i)\}\{C(a)C(a-3)C(a-4)\}^{-1} \\ & \times \{C^2(a-1) - C(a)C(a-2)\}^{-2} - 3, \quad a > 6.\end{aligned}$$

Proof: To prove (iv), note that the Pearson's moment coefficient of skewness is given by

$$\gamma_1(X) = \frac{m_{X,3}}{\sigma_X^3},$$

where $m_{X,3} = \mathbb{E}[X - \mathbb{E}(X)]^3$, i.e., the 3th central moment of the random variable X and σ_X is the standard deviation of X. Thus, using the properties of expectation and (3.4), it can be shown that

$$m_{X,3} = \frac{C(a)[C^3(a-1)C(a-2) - 3C(a)C^2(a-1)C(a-3) + 2C^2(a)C(a-2)C(a-3)]}{C^3(a-1)C(a-2)C(a-3)}$$

Using (3.5), $[Var(X)]^{3/2} = \sigma_X^3 = \frac{C^{3/2}(a)[C^2(a-1) - C(a)C(a-2)]^{3/2}}{C^{3/2}(a-2)C^3(a-1)}$, so that the coefficient can be calculated.

To prove (v), note that the excess kurtosis is calculated by $K(X) = m_{X,4}/\sigma_X^4 - 3$. The computation of the fourth central moment is analogous to that of the coefficient of skewness, and the denominator σ_X^4 is simply the squared variance given in (3.5). □

Figure 2 shows how the expectation and variance behave as the parameter a changes. Note that as the parameter a increases, both the expectation and variance of the standard floor distribution decrease, however the variance decreases more rapidly.

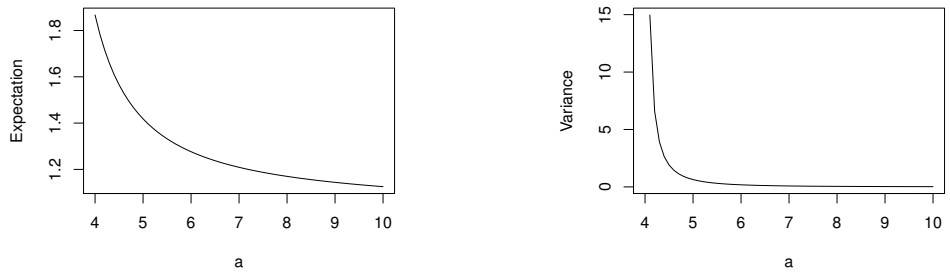


Figure 2: Expectation and variance for different values of a.

Figure 3 presents the behaviour of the skewness and kurtosis coefficients as the value of a changes. Note that the value of $\gamma_1(X)$ will be always greater than *zero*, which indicates that the floor distribution is right-skewed for any value of $a > 5$.

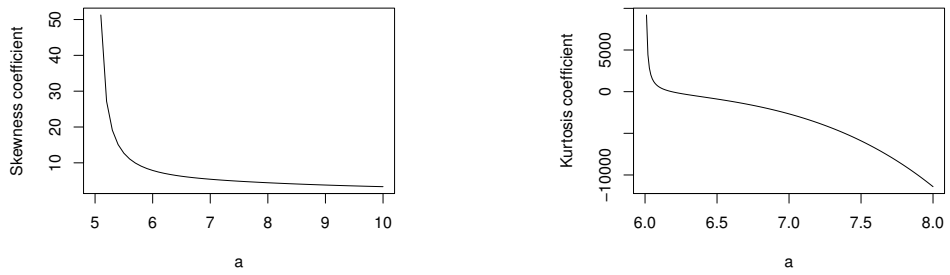


Figure 3: Skewness and kurtosis coefficients for different values of a.

3.6. Summaries for the floor distribution with location and scale

Proposition 3.2. *If $Z = \sigma X + \mu$ is a random variable following the floor distribution with location and scale parameters given by μ and σ, respectively, then*

(**i**) *The expectation of Z, given straightforwardly by (3.4), and using properties of expectation is given by* $\mathbb{E}(Z) = \sigma\mathbb{E}(X) + \mu$ $a \geq 3, \mu \in \mathbb{R}, \sigma > 0$.

(**ii**) *Variance. Using (3.4), and properties of variance, we have*

$$Var(Z) = \sigma^2 Var(X) = \sigma^2 \frac{C(a)[C^2(a-1) - C(a)C(a-2)]}{C(a-2)C^2(a-1)}, \ a > 4. \tag{3.6}$$

(**iii**) *The moment of order r of the variable Z can be obtained using the binomial theorem, i.e.,* $\mathbb{E}(Z^r) = \sum_{k=0}^{r} \binom{r}{k} \mu^{r-k}\sigma^k \mathbb{E}(X^k)$.

(**iv**) *Quantiles can be obtained by numerically inverting the cumulative distribution function.*

(**v**) *The coefficient of skewness is the same as of the standard floor distribution, i.e.,* $\gamma_1(Z) = \gamma_1(X)$.

(**vi**) *The excess kurtosis coefficient is the same as of the standard floor distribution, i.e.,* $K(Z) = K(X)$.

Proof: To see that the coefficient of skewness and the excess kurtosis are the same for the standard floor distribution and the floor distribution with location and scale parameter, note that $\gamma_1(Z) = \frac{m_{Z,3}}{\sigma_Z^3}$, where

$$\begin{aligned} m_{Z,3} &= \mathbb{E}\{(Z - \mathbb{E}(Z))^3\} = \mathbb{E}\{(\sigma X + \mu - (\sigma\mathbb{E}(X) + \mu))^3\} \\ &= \mathbb{E}\{\sigma^3(X - \mathbb{E}(X))^3\} = \sigma^3\mathbb{E}\{(X - \mathbb{E}(X))^3\} = \sigma^3 m_{X,3} \end{aligned}$$

and $\sigma_Z^3 = (\sigma_Z^2)^{\frac{3}{2}} = \{\sigma^2 Var(X)\}^{\frac{3}{2}} = (\sigma\sigma_X)^3$. Thus

$$\gamma_1(Z) = \frac{m_{Z,3}}{\sigma_Z^3} = \frac{\sigma^3 m_{X,3}}{(\sigma\sigma_X)^3} = \frac{m_{X,3}}{\sigma_X^3} = \gamma_1(X)$$

with similar computation, it can be shown that $K(Z) = K(X)$. □

3.7. Random values of the floor distribution

Through the acceptance-rejection method, values of the floor distribution are generated (Kronmal & Peterson, 1981) using the R software. We considered the exponential distribution with location parameter equal to 1 as the proposed distribution, so that the support of both the exponential and floor distributions are the same.

Consider the density of the floor distribution as $f(x)$ and the density of the exponential distribution as $g(y)$. The acceptance-rejection algorithm used to generate random values of a standard floor distribution with parameter a can be explicitly written as follows:

1. Generate a random value of the exponential distribution with two parameters, choosing λ so that the floor and the exponential distribution have the same expectation and fixing the location parameter as equal to 1. Call this value y.
2. Generate a random value of the uniform distribution, u.
3. If $u \leq \frac{f(y)}{cg(y)}$, set x=y (accept). Otherwise go back to step 1 to generate a new value.

In the third step, we assume that the ratio between $f(x)$ and $g(x)$ is bounded by c, a constant greater than zero.

Figure 4 shows a histogram with the distribution of 10,000 random values, generated through the acceptance-rejection method for $a = 9$, with the real density over the histogram.

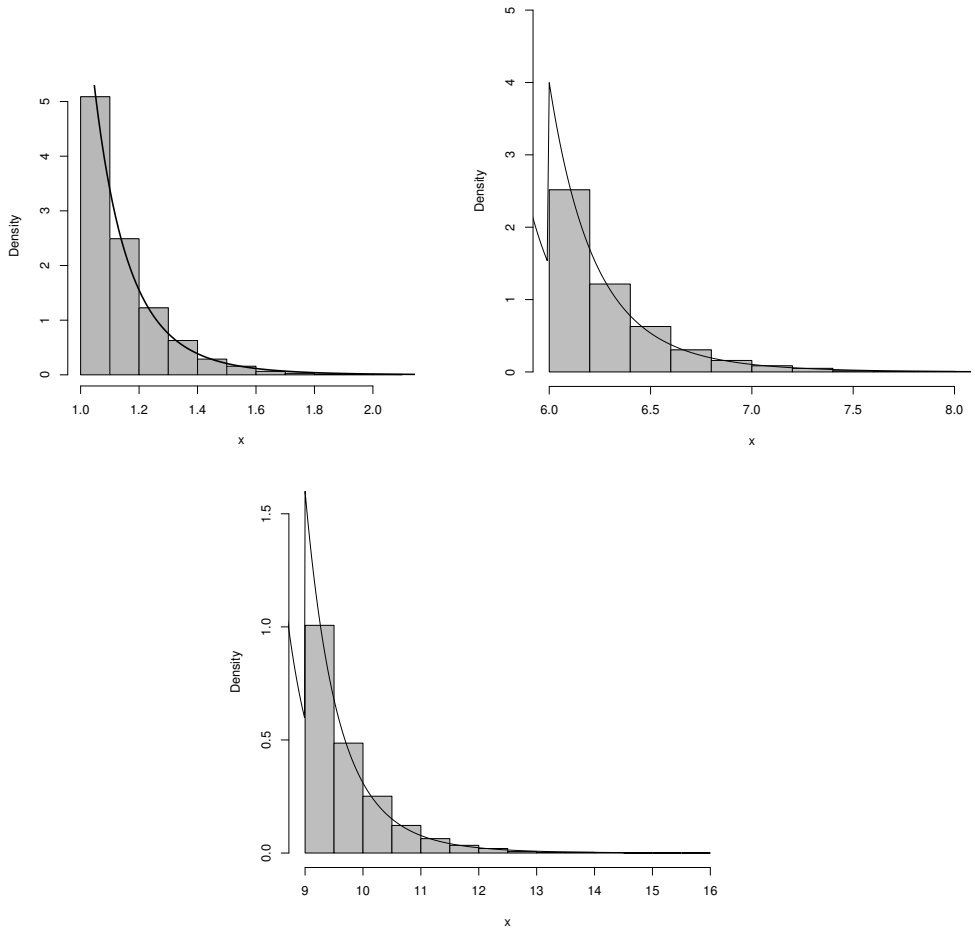


Figure 4: Histograms of generated values with $a = 9$, $\mu = 0$, $\sigma = 1$; $a = 9$, $\mu = 4$, $\sigma = 2$; $a = 9$, $\mu = 4$, $\sigma = 5$, respectively.

4. SIMULATION STUDY

A simulation study was performed to compare the floor distribution to the exponential distribution, under different sample sizes and proportion of outliers. We used samples of sizes $20, 80$ and 100, and for each sample size we considered three proportions of outliers: $0.05, 0.10$ and 0.15.

The Figure 5 shows how the posterior estimation for the location parameter behaved as the outlying observations increased. The floor distribution is represented by the dashed line and the exponential distribution is represented by the continuous line. As the Figure 5 shows, the estimation of the location parameter under the floor distribution was not so affected as under the exponential distribution.

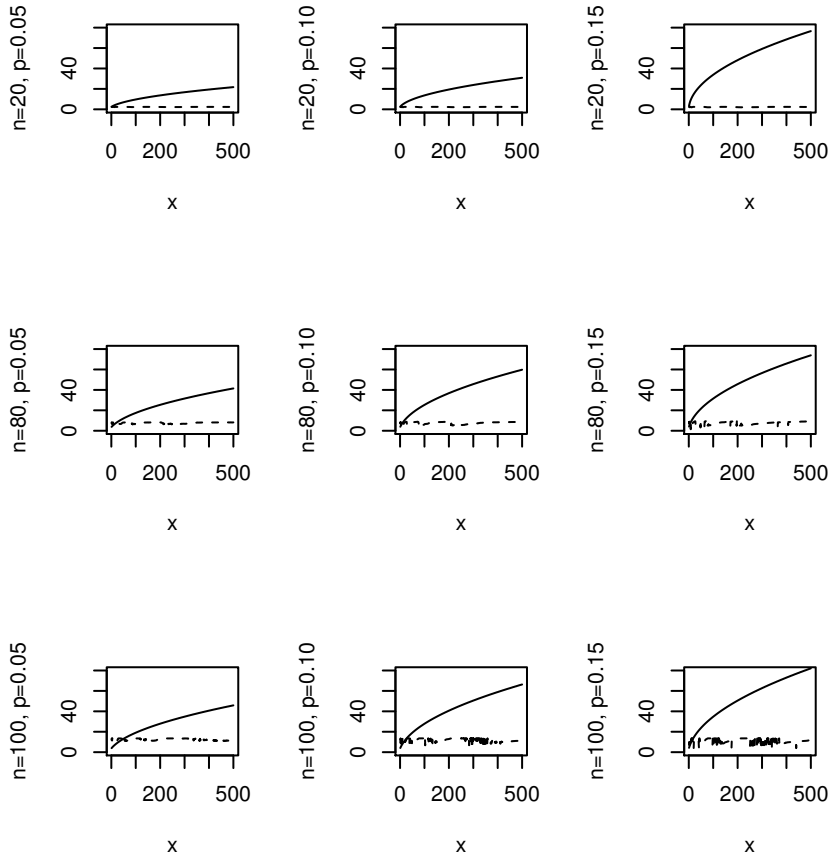


Figure 5: Comparison of location parameter posterior estimation for the floor and exponential distributions under different sample sizes and proportions of outliers.

5. APPLICATIONS — BAYESIAN ROBUSTNESS

In this section, we illustrate how the floor distribution can resolve conflict of information by rejecting the outlying information. We compare the behaviour of the posterior estimates under a floor distribution and exponential distribution models. The usual procedure to assess robustness in a Bayesian model is to make one (or a few) observations in the data to tend to infinity, and check how the posterior estimates behaves. Thus, modelling accordingly to conditions (i)–(iv) which lead to the result (2.7), the posterior distribution will automatically reject the conflicting observation.

We use the data from Kapur & Lamberson (1977, p. 240), which refers to X: the number cycles to failure (in ten thousands) for 20 heater switches

subject to an overload voltage. Following the same methodology used in Andrade & Omey (2016), the data are modelled with two different densities with scale parameter: exponential, with density given by $f(x|\theta) = \theta^{-1}e^{-x/\theta}, x > 0$ and floor, with density given by

$$f(x|\theta) = \frac{C(a)}{\theta}\left(\frac{x}{\theta}\right)^{-a} e^{\log\lfloor x/\theta \rfloor},$$

for $x > \theta$. The conditions established by Andrade & Omey (2016) consider location and scale parameter structures separately, thus a more complex structure, involving location and scale parameters, will change their conditions, hence more investigation is required in order to assess the behaviour of the posterior quantities in the location-scale parameter case, likewise to Andrade & O'Hagan (2011).

Tahir & Saleem (2011) considered a elicited prior density is which is quite informative, in the sense that the prior variance is relatively small. Thus, $\theta \overset{d}{\sim}$ Gama$(8.9936, 21.5698)$, which we will also use for both the floor and exponential models.

We used the package `OpenBugs` with zeros trick, since the floor is a new distribution. As a result, Figure 6 was created by simulating the posterior distribution for each model, as one of the observations tends to infinity. In the case of the exponential distribution, as the outlier becomes distant from the other observations, the posterior mean is affected by the outlying information. This does not happen with the floor distribution, and thus we can see from this example that the floor distribution is robust to outliers.

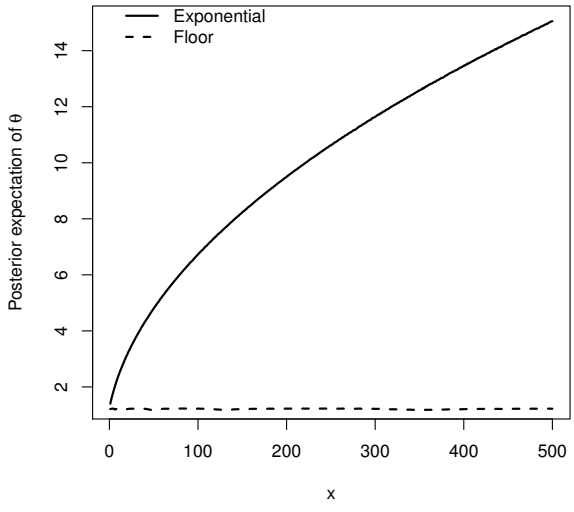


Figure 6: Posterior estimates for θ.

6. DISCUSSION

In Bayesian context, robustness modelling is becoming of high interest, mainly to address problems due to misspecification of the model. In fact, in a rigorous modelling process, a researcher may change the model after detecting conflicting information such as outliers, untrusted prior information, etc. Therefore, it is important to know about the properties of the heavy-tailed distributions in order to model conveniently to resolve such conflicts. However, the large range of heavy-tailed distributions leads to a great variety of behaviours of the posterior distribution in the presence of conflict, some heavy-tailed distribution will yield robust models only for the location parameter, whereas other classes of such distributions will resolve conflicts in both location and scale parameters. In addition, different classes can lead to different ways to resolve the conflict. For instance, as pointed out by Andrade & O'Hagan (2006), in the scale parameter case, the regularly varying distributions will allow only a partial rejection of the conflicting information, whereas the class proposed by Desgagné (2013, 2015) achieve full rejection. In this work, we follow the proposal of Andrade & Omey (2016), in which uses the ORV class, which also lead to complete rejection the outlier, however the ORV class is much more intuitive and easy to work than that proposed by Desgagné (2013).

On the other hand, there are very few ORV distributions in the literature, further work should propose new ORV distributions. The floor distribution is an alternative to the exponential distribution, and is an example of how the tails can oscillate leading to a new sort of heavy-tailed distributions, which has direct applications in Bayesian Robustness.

ACKNOWLEDGMENTS

The first author is grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the support in this work. We thank the referees for the very useful suggestions which improved the text.

REFERENCES

[1] Andrade, J.A.A.; Dorea, C.C.Y. and Otiniano, C.E.G. (2013). On the Robustness of Bayesian Modelling of Location and Scale Structures Using Heavy-Tailed Distributions, *Communication in Statistics: Theory and Methods*, **42**(8), 1502–1514.

[2] Andrade, J.A.A. and O'Hagan, A. (2006). Bayesian robustness modelling using regularly varying distributions, *Bayesian Analysis*, **1**(1), 169–188.

[3] Andrade, J.A.A. and O'Hagan, A. (2011). Bayesian Robustness Modelling of Location and Scale Parameters, *Scandinavian Journal of Statistics*, **38**, 691–711.

[4] Andrade, J.A.A. and Omey, E. (2013). Modelling conflicting information using subexponential distributions and related classes, *Annals of the Institute of Statistics and Mathematics*, **65**, 491–511.

[5] Andrade, J.A.A. and Omey, E. (2016). Resolution of conflict of information using O-regularly varying functions, Submitted to *Statistica Neerlandica*.

[6] Avakumović, V.G. (1936). Über einen O-inversionssatz, *Bulletin International de l'Academie Yougoslave des Sciences et des Beaux-Arts*, **29**(30), 107–117.

[7] Bingham, N.H.; Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*, "Encyclopedia of Mathematics and Its Applications", Cambridge: Cambridge University Press.

[8] Dawid, A.P. (1973). Posterior Expectations for Large Observations, *Biometrika*, **60**(3), 664–667.

[9] De Haan, L. (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*, Mathematical Centre Tracts 32, Mathematisch Centrum, Amsterdam.

[10] Desgagné, A. (2013). Full Robustness in Bayesian Modelling of a Scale Parameter, *Bayesian Analysis*, **8**(1), 187–220.

[11] Desgagné, A. (2015). Robustness to outliers in location-scale parameter model using log-regularly varying distributions, *The Annals of Statistics*, **43**(4), 1568–1595.

[12] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, second edition, John Wiley & Sons.

[13] Kapur, K.C. and Lamberson, L.R. (1977). *Reliability in Engineering Design*, New York: John Wiley & Sons, Inc.

[14] Karamata, J. (1930). Sur un mode de croissance régulière des fonctions, *Mathematica (Cluj)*, **4**, 38–53.

[15] Kronmal, R.A. and Peterson Jr, A.V. (1981). A Variant of the Acceptance-Rejection Method for Computer Generation of Random Variables, *Journal of the American Statistical Association*, **76**(374), 446–451.

[16] Landau, E. (1911). Sur les valeurs moyennes de certaines fonctions arithmetiques, *Bulletin de l'Academie Royale de Belgique*, 443–472.

[17] Lindley, D.V. (1968). The choice of variables in multiple regression (with discussion), *Journal of the Royal Statistical Society, Series B*, **30**(1), 31–66.

[18] O'Hagan, A. (1979). On outlier rejection phenomena in Bayes inference, *Journal of the Royal Statistical Society, Series B*, **41**(3), 358–367.

[19] O'Hagan, A. (1988). Modelling with Heavy Tails, *Bayesian Statistics*, **3**, 345–359.

[20] O'Hagan, A. (1990). Outliers and Credence for Location Parameter Inference, *Journal of the American Statistical Association*, **85**(40), 172–176.

[21] O'Hagan, A. and Le, H. (1994). Conflicting information and a class of bivariate heavy-tailed distributions. In *Aspects of Uncertainty: a Tribute to D.V. Lindley*, (P.R. Freeman and A.F.M. Smith, Eds.), New York: Wiley, 311–327.

[22] O'Hagan, A. and Pericchi, L. (2012). Bayesian heavy-tailed models and conflict resolution: a review, *Brazilian Journal of Probability and Statistics*, **26**(4), 372–401.

[23] Pólya, G. (1917). Über eine neue Weise, bestimmte Integrale in der analytischen Zahlentheorie zu gebrauchen, *Nachrichten Göttinger*, **1917**(1), 149–159.

[24] Seneta, E. (1976). *Functions of regular variation*, Lecture Notes in Mathematics Vol. 506, New York: Springer.

[25] Tahir, M. and Saleem, M. (2011). A comparison of priors for the parameter of the time-to-failure model, *Pakistan Journal of Science*, **63**(1), 49–52.

[26] Valiron, G. (1913). Sur les fonctions entières d'order fini, et d'ordre nul, et en particulière les fonctions à correspondence régulière, *Annales de la Faculté des sciences de Toulouse*, **3**(5), 117–257.

REVSTAT – Statistical Journal
Volume 16, Number 4, October 2018, 463–474

PARAMETRIC TESTS OF PERFECT JUDGMENT RANKING BASED ON ORDERED RANKED SET SAMPLES

Authors: Ehsan Zamanzade
– Department of Statistics, University of Isfahan, Isfahan 81746-73441, Iran
E.Zamanzade@sci.ui.ac.ir ; Ehsanzamanzadeh@yahoo.com

Michael Vock
– Institute of Mathematical Statistics and Actuarial Science, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland
michael.vock@stat.unibe.ch

Received: January 2016 Revised: November 2016 Accepted: November 2016

Abstract:

- We develop parametric and location-scale free tests of perfect judgment ranking based on ordered ranked set samples. The tests are based on the differences between the elements of the ordered ranked set samples and those of the original ranked set samples. We compare our proposed tests with the best existing tests of perfect judgment ranking in the literature by using Monte Carlo simulation. Our simulation results show that the proposed tests behave favorably in comparison with their leading competitors, especially under the fraction of neighbor rankings model. In comparison to the nonparametric competitors, the proposed tests have the advantage of not needing randomization to attain a specific size.

Key-Words:

- *ordered ranked set samples; perfect ranking; test.*

AMS Subject Classification:

- 62D05, 62F03.

1. INTRODUCTION

When measuring variables of interest is expensive or time-consuming, but ranking them in small groups without actual measurement is easy and convenient, ranked set sampling (RSS) can be regarded as an efficient technique for collecting more informative samples and therefore having more reliable inferences. This sampling technique, which was firstly introduced by McIntyre (1952, 2005), can be applied in both balanced and unbalanced strategies. In the balanced case, the researcher first draws k random samples of size k and orders them based on his personal judgment (not actual measurement). Then, for $i = 1, ..., k$, he actually measures the i^{th} judgment ordered observation from the i^{th} sample. Finally, he repeats this procedure n times (cycles) in order to draw a sample of size kn from a Balanced Ranked Set Sampling (BRSS) scheme. In Unbalanced Ranked Set Sampling (UBRSS), the numbers of i^{th} judgment ordered observations are not necessarily the same anymore. A comprehensive review of works on RSS including a comprehensive list of references can be found in Wolfe (2012).

Although many researchers have shown that a ranked set sample may allow for more reliable inferences than a simple random sample of the same size, this reliability decreases as errors in ranking observations based on personal judgment occur. Frey *et al.* (2007) have exemplified how the ranking error can invalidate the method of inference in both parametric and nonparametric cases. Therefore, it seems to be vital to develop tests for assessing the assumption of perfect judgment ranking for both parametric and nonparametric cases. Surprisingly, this has not been done up to quite recently. Frey *et al.* (2007) and Li and Balakrishnan (2008) independently proposed some nonparametric tests of perfect judgment ranking, followed by Vock and Balakrishnan (2011), Zamanzade *et al.* (2012), Vock and Balakrishnan (2013), Frey and Wang (2013), and Zamanzade *et al.* (2014).

This paper is organized as follows: In Section 2, we propose our tests of perfect judgment ranking for one cycle, then, in the next section, we generalize them to the multi-cycle case. In Section 4, we compare our proposed tests with their leading competitors in the literature. Conclusions and some final remarks are provided in Section 5.

2. INTRODUCTION OF TESTS STATISTICS

Let $X_{[1]}, ..., X_{[k]}$ be a sample of size k from BRSS with one cycle, where $X_{[i]}$ $(i = 1, ..., k)$ is the i^{th} judgment ordered observation from the i^{th} sample, which is actually measured. It should be noted that the $X_{[i]}$'s are independent from each other and follow the distribution of an i^{th} order statistic if the assumption of perfect judgment ranking is completely satisfied. Furthermore, due to

the independence of the $X_{[i]}$'s, $P\left(X_{[i]} < X_{[j]}\right) < 1$ for $i < j$ and $i, j \in \{1, ..., k\}$, and this probability decreases as the judgment ranking becomes more and more unreliable. So intuitively, it is expected that the two vectors $\left(X_{[1]}, ..., X_{[k]}\right)$ and $\left(Z_{(1)}, ..., Z_{(k)}\right)$ are close to each other provided that the assumption of perfect judgment ranking is completely satisfied, where $\left(Z_{(1)}, ..., Z_{(k)}\right)$ is the vector of Ordered Ranked Set Samples (ORSS) which is obtained by putting the values of $\left(X_{[1]}, ..., X_{[k]}\right)$ in order. Therefore if the underlying distribution of population is completely known, then the following tests can be proposed for assessing the assumption of perfect judgment ranking:

$$TA = \sum_{i=1}^{k} \frac{|d_i|}{E|d_i|};$$

$$TS = \sum_{i=1}^{k} \frac{d_i^2}{Ed_i^2};$$

where $d_i = X_{[i]} - Z_{(i)}$, and $E(.)$ is the expectation operator which is taken under the assumption of perfect judgment ranking.

Intuitively, large values of TA, TS are a symptom of violation of the assumption of perfect judgment ranking and therefore this assumption should be rejected for large enough values of TA, TS.

If the underlying distribution of the population belongs to a location-scale family, then the above test statistics can be simplified as follows:

$$TA = \sum_{i=1}^{k} \frac{|d_i|}{\sigma E_{\mu=0,\sigma=1}|d_i|};$$

$$TS = \sum_{i=1}^{k} \frac{d_i^2}{\sigma^2 E_{\mu=0,\sigma=1} d_i^2};$$

where μ, σ are location and scale parameters, respectively.

Obviously, the above test statistics are location-free, and they will be scale-free if an equivariant estimator is used for the estimation of σ.

3. EXTENSION OF THE PROPOSED TESTS TO THE MULTI-CYCLE CASE

Although several methods have been proposed in the literature for extending tests of perfect judgment ranking from the one-cycle to the multi-cycle case, Zamanzade *et al.* (2012)'s simulation study has shown that their permutation-based technique provides good results under many scenarios. So we use their method to extend our tests to the multi-cycle case.

Suppose that $\left(X_{[i]j}\right)_{i\leq k, j\leq n}$ is a sample of size of kn, which is drawn by an n-cycle BRSS scheme, where $X_{[i]j}$ is the i^{th} judgment ordered observation from the j^{th} cycle $(i = 1, ..., k; j = 1, ..., n)$. Since all observations are mutually independent and the observations in each column are also identically distributed, it is expected that the vector of ordered observations in each row should be close to the unordered row vector if we permute observations in each column, provided that the assumption of perfect judgment ranking is fully satisfied. In other words, under the assumption of perfect judgment ranking, the vector $\left(X_{[1]l_1}, X_{[2]l_2}, X_{[3]l_3}, ..., X_{[k]l_k}\right)$ and the ordered vector of this vector, which is denoted here by $\left(Z_{(1)l_1l_2...l_k}, Z_{(2)l_1l_2...l_k}, Z_{(3)l_1l_2...l_k}, ..., Z_{(k)l_1l_2...l_k}\right)$, should be close to each other for all $(l_1, l_2, l_3, ..., l_k) \in \{1, 2, 3, ..., n\}^k$.

Based on the above arguments, TA, TS can be extended to the multi-cycle case as follows:

$$TPA = \sum_{i=1}^{n^k} TA_i,$$

$$TPS = \sum_{i=1}^{n^k} TS_i,$$

where TA_i, TS_i are the values of TA and TS, respectively, for the i^{th} sample out of all n^k samples of the form $\left(X_{[1]l_1}, X_{[2]l_2}, X_{[3]l_3}, ..., X_{[k]l_k}\right)$, $(l_1, l_2, l_3, ..., l_k) \in \{1, 2, 3, ..., n\}^k$.
We reject the hypothesis of perfect judgment ranking for large enough values of TPA and TPS.

The calculation of TPA or TPS based on all n^k samples of the form mentioned above is too time-consuming for practical application except for very small values of k and n. We therefore propose a less intuitive, but more efficient way of computing these statistics. R-code for the computation of TPA and TPS using the following method is available on request from the authors.

For n cycles, with $E_i = E_{\mu=0,\sigma=1}|d_i|$, TPA can be written as

$$\begin{aligned} TPA &= \sum_{(l_1,...,l_k)\in\{1,...,n\}^k} \sum_{i=1}^{k} \frac{\left|X_{[i]l_i} - Z_{(i)l_1l_2,...,l_k}\right|}{\sigma E_i} \\ &= \sum_{i=1}^{k} \frac{1}{\sigma E_i} \sum_{l_i=1}^{n} \sum_{(l_1,...,l_{i-1},l_{i+1},...,l_k)\in\{1,...,n\}^{k-1}} \left|X_{[i]l_i} - Z_{(i)l_1l_2...l_k}\right| \\ &= \sum_{i=1}^{k} \frac{1}{\sigma E_i} \sum_{l_i=1}^{n} \sum_{j=1}^{k} \sum_{h=1}^{n} m\left(i, l_i, j, h\right) \left|X_{[i]l_i} - X_{[j]h}\right| \end{aligned}$$

where $m\left(i, l_i, j, h\right)$ is the number of vectors $(l_1, ..., l_{i-1}, l_{i+1}, ..., l_k) \in \{1, ..., n\}^{k-1}$ such that the i^{th} order statistic from $X_{[1]l_1}, ..., X_{[k]l_k}$ is the j^{th} judgment ordered

observation from the h^{th} cycle. (A similar representation applies to TPS.) Since for each judgment order rank, only one cycle is used, this implies that $l_j = h$, and $m\,(i, l_i, j, h)$ is actually the number of vectors $(l_q)_{q\in\{1,\ldots,k\}\setminus\{i,j\}} \in \{1, ..., n\}^{k-2}$ such that the i^{th} order statistic from $X_{[1]l_1}, ..., X_{[k]l_k}$ is the j^{th} judgment ordered observation from the h^{th} cycle.

In the following, we assume that there are no ties. Since (1) $m\,(i, l_i, j, h)$ is 0 if $i = j$ and $l_i \neq h$ and (2) $X_{[i]l_i} - X_{[j]h} = 0$ if $i = j$ and $l_i = h$, j can be assumed to be different from i:

$$TPA = \sum_{i=1}^{k} \frac{1}{\sigma E_i} \sum_{l_i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{k} \sum_{h=1}^{n} m\,(i, l_i, j, h) \left| X_{[i]l_i} - X_{[j]h} \right|.$$

We therefore only need the values of $m\,(i, l_i, j, h)$ for $i \neq j$. For $i = 1, ..., k$, let $a\,(i, j, h)$ be the number of observations in the i^{th} judgment-order stratum that are smaller than $X_{[j]h}$,

$$a\,(i, j, h) = \#\left\{l \in \{1, ..., n\} : X_{[i]l} < X_{[j]h}\right\}.$$

Then, by using the fact that exactly $i - 1$ observations from $X_{[1]l_1}, ..., X_{[k]l_k}$ (of which $X_{[i]l_i}$ may be one or not) have to be smaller than the i^{th} order statistic from $X_{[1]l_1}, ..., X_{[k]l_k}$,

$$\begin{aligned} m\,(i, l_i, j, h) \;=\; & \#\Bigg\{ (l_q)_{q\in\{1,\ldots,k\}\setminus\{i,j\}} \in \{1, ..., n\}^{k-2} : \\ & \qquad \sum_{q\in\{1,\ldots,k\}\setminus\{i,j\}} I(X_{[q]l_q} < X_{[j]h}) = i - 1 - I(X_{[i]l_i} < X_{[j]h}) \Bigg\} \\ = \; & \sum_{\substack{Q\subset\{1,\ldots,k\}\setminus\{i,j\} \\ \#Q = i-1-I(X_{[i]l_i}<X_{[j]h})}} \prod_{q\in Q} a\,(q, j, h) \prod_{q\notin Q} (n - a\,(q, j, h)) \end{aligned}$$

where $I(.)$ is the indicator function.

The $a\,(i, j, h)$'s can be calculated efficiently by going through all kn observed values $X_{[j]h}$ in increasing order and using the fact that $a\,(i, j^*, h^*) = 0$ (for $i = 1, ..., k$) if $X_{[j^*]h^*}$ is the smallest value from the sample, as well as the following recursions, where $X_{[j]h}$ and $X_{[j']h'}$ are assumed to be two successive values of the ordered sample:

$$a(i, j', h') = \begin{cases} a(i, j, h) + 1 & \text{if } i = j, \\ a(i, j, h) & \text{if } i \neq j. \end{cases}$$

E.g., for $k = n = 8$, this approach for the computation of TPA and TPS resulted in a reduction of the computation time by a factor of approximately 1200 compared to the original algorithm.

4. POWER COMPARISON

In this section, we compare the power of our proposed tests with their leading competitors in the literature under the assumption that the parent distribution is normal. The competing tests considered here are as follows:

- Nonparametric test based on W^* developed by Frey *et al.* (2007), which rejects the hypothesis of perfect judgment ranking when $W^* = \sum_{i=1}^{k}\sum_{j=1}^{n} iR_{[i]j}$ is too small, where $R_{[i]j}$ is the rank of $X_{[i]j}$ among all kn observations.
- Nonparametric test based on the null probability (NP) developed by Frey *et al.* (2007), which rejects the hypothesis of perfect judgment ranking when the null probability of observing set rank $\left\{R_{[i]j}\right\}$ is too small.
- Nonparametric test based on J developed by Vock and Balakrishnan (2011), which rejects the hypothesis of perfect judgment ranking when $J = \sum_{h=1}^{n}\sum_{l=1}^{n}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} I\left(X_{[i]l} > X_{[j]h}\right)$ is too large.
- Nonparametric test based on PA developed by Zamanzade *et al.* (2012), which rejects the hypothesis of perfect judgment ranking when $PA = \sum_{h=1}^{n^k}\sum_{i=1}^{k}\left|R^*_{[i]h} - i\right|$ is too large, where $R^*_{[i]h}$ is the rank of the i^{th} judgment ordered observation in the h^{th} permuted sample introduced in Section 3.
- Parametric test based on D developed by Zamanzade *et al.* (2014), which rejects the hypothesis of perfect judgment ranking when $D = \sum_{h=1}^{n}\sum_{l=1}^{n}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\frac{\left(X_{[i]l}-X_{[j]h}\right)I\left(X_{[i]l}>X_{[j]h}\right)}{E_{\mu=0,\sigma=1}\left(\left(X_{[i]l}-X_{[j]h}\right)I\left(X_{[i]l}>X_{[j]h}\right)\right)}$ is too large, where μ, σ are location and scale parameters, respectively.
- Most powerful rank test (MP) developed by Frey and Wang (2013). In this test, it is assumed that the alternative hypothesis of perfect judgment ranking is fully specified, i.e. the underlying distribution of the population, the scenario of imperfect ranking, and the fraction of imperfect ranking are all completely known. Then the null hypothesis is rejected when $r = \frac{P_{H_1}(W_1<W_2<\cdots<W_N)}{P_{H_0}(W_1<W_2<\cdots<W_N)}$ is too large, where W_i has the same distribution as the in-set rank of the observation with rank i among all the $N = kn$ measured values.

We assume that the parent distribution is normal with unknown mean μ and unknown variance σ^2. We don't need to estimate the parameter μ because the proposed tests are location free. The parameter σ is estimated by $\hat{\sigma} = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}\left(X_{[i]} - \bar{X}\right)^2}$ for $n = 1$ due to Stokes (1980) and by $\hat{\sigma} =$

$\sqrt{\left(\frac{1}{nk}+\frac{1}{nk^2(n-1)}\right)\sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{[i]j}-\bar{X}_{[i].}\right)^2+\frac{1}{k}\sum_{i=1}^{k}\left(\bar{X}_{[i].}-\bar{X}\right)^2}$ for $n>1$ as proposed by MacEachern *et al.* (2002) and Perron and Sinha (2004), where $\bar{X}_{[i].}$ is the mean of the observations with judgment rank i. Obviously, these estimators of σ are equivariant and the resulting test statistics (denoted by $\widehat{TPA}, \widehat{TPS}$ and $\widehat{D}$) are scale invariant. Therefore, the critical values and powers of tests based on $\widehat{TPA}, \widehat{TPS}$ and $\widehat{D}$ don't depend on the unknown parameters μ and σ. The expected values $E\left(d_i^2\right)$ and $E\left|d_i\right|$ and critical values of the tests based on $\widehat{TPA}$ and $\widehat{TPS}$ under the assumption of normality are available on request from the authors.

In our simulation study, the comparisons are done at a significance level of $\alpha=0.05$. However, due to the discreteness of the distribution of the nonparametric test statistics, it is not possible to attain an exact size of $\alpha=0.05$ without randomizing. Therefore, we have used the randomized versions of those tests to make all comparisons at size $\alpha=0.05$. For example, for $n=1$ and $k=5$, under the assumption of perfect ranking, $J\geq 4$ with null probability 0.03345, and $J\geq 3$ with null probability 0.12687. Thus in order to attain the significance level $\alpha=0.05$ in a randomized test based on J, H_0 is rejected with probability one if $J\geq 4$ and with probability $\frac{0.05-0.03345}{0.12687-0.03345}=0.177$ if $J=3$.

We have used two different scenarios of imperfect ranking, which have been used by many researchers in the literature. The first scenario is the bivariate normal model, due to Dell and Clutter (1972), in which the variable of interest X is ordered by using a concomitant variable Y, where (X,Y) has a bivariate normal distribution with correlation coefficient λ.

The second scenario is that of a fraction of neighbor rankings, developed by Vock and Balakrishnan (2011), in which the i^{th} judgment ordered observation is either ranked perfectly with probability λ, or is confused with the $(i+1)^{\text{th}}$ or $(i-1)^{\text{th}}$ ordered observation, both with probability $\frac{\lambda}{2}$, therefore the distribution of the i^{th} judgment ordered observation under this scenario is $F_{[i]}=\frac{\lambda}{2}F_{(i-1)}+(1-\lambda)F_{(i)}+\frac{\lambda}{2}F_{(i+1)}$, where $F_{(0)}=F_{(1)}, F_{(k+1)}=F_{(k)}$. This imperfect ranking model could arise when the ranking process is done by using personal judgment of an expert ranker, so he may confuse the true order statistic with an adjacent one.

For power comparisons, we have extended Tables 3 and 6 of Frey and Wang (2013) to all tests introduced above and larger values of (n,k) by using Monte Carlo simulation with 100,000 repetitions. The simulation results are presented in Tables 1–2. It should be noted that in the following tables the powers of the tests based on MP, NP, W^*, J for $(n,k)=(8,2),(4,3),(2,4),(1,5)$ are directly reported from Tables 3 and 6 of Frey and Wang (2013). Furthermore, we haven't estimated the power of the MP test for $(n,k)=(4,5),(5,4)$, since this test is only applicable for small sample sizes and small set sizes because of its computational limitations.

Table 1 gives power results for the bivariate normal model. It is apparent from this table that although the NP test is the most powerful among the non-parametric tests, the powers of the test based on W^* are quite close to it. On the other hand, the proposed tests and the test based on $\widehat{D}$ have the best powers in this scenario, and the differences between their powers are not considerable.

Table 1: Power estimates of different level 0.05 tests, under the concomitant model with correlation coefficient λ.

k	n	λ	$\widehat{TPS}$ (new)	$\widehat{TPA}$ (new)	$\widehat{D}$ (Z., *et al.* 2014)	MP (Frey and Wang, 2013)	NP (Frey *et al.*, 2007)	W^* (Frey *et al.*, 2007)	J (V. and Balakr., 2011)	PA (Z. *et al.*, 2012)
2	8	0.9	.1047	.1042	.1060	.1019	.1018	.1000	.1000	.1000
		0.8	.1880	.1815	.1815	.1722	.1720	.1676	.1676	.1676
		0.7	.2715	.2715	.2710	.2566	.2563	.2490	.2490	.2490
		0.6	.3679	.3642	.3671	.3496	.3493	.3394	.3394	.3394
		0.5	.4656	.4694	.4678	.4458	.4456	.4337	.4337	.4337
3	4	0.9	.1265	.1384	.1302	.1317	.1316	.1294	.1289	.1271
		0.8	.2400	.2465	.2487	.2401	.2400	.2346	.2336	.2344
		0.7	.3599	.3729	.3628	.3596	.3594	.3509	.3496	.3515
		0.6	.4678	.4858	.4947	.4777	.4776	.4669	.4655	.4625
		0.5	.5891	.6002	.5959	.5867	.5866	.5748	.5783	.5734
4	2	0.9	.1405	.1425	.1509	.1491	.1420	.1403	.1372	.1398
		0.8	.2599	.2556	.2663	.2555	.2553	.2511	.2448	.2477
		0.7	.3678	.3894	.3882	.3720	.3718	.3653	.3568	.3613
		0.6	.4776	.4811	.4984	.4819	.4818	.4737	.4640	.4697
		0.5	.5763	.5808	.6006	.5808	.5806	.5718	.5617	.5650
5	1	0.9	.1367	.1365	.1485	.1366	.1363	.1358	.1283	.1254
		0.8	.2413	.2355	.2582	.2335	.2332	.2316	.2174	.2113
		0.7	.3428	.3335	.3658	.3287	.3287	.3260	.3071	.2960
		0.6	.4345	.4242	.4615	.4182	.4182	.4147	.3929	.3779
		0.5	.5213	.5103	.5456	.5003	.5003	.4962	.4731	.4532
5	4	0.9	.2697	.2681	.2843	—	.2969	.2866	.2845	.2846
		0.8	.5471	.5437	.5538	—	.5811	.5649	.5625	.5632
		0.7	.7590	.7577	.7541	—	.7820	.7665	.7651	.7648
		0.6	.8841	.8841	.8733	—	.8934	.8818	.8819	.8816
		0.5	.9494	.9395	.9380	—	.9506	.9436	.9434	.9436
4	5	0.9	.2247	.2286	.2181	—	.2261	.2224	.2231	.2182
		0.8	.4508	.4664	.4387	—	.4618	.4520	.4523	.4462
		0.7	.6446	.6633	.6354	—	.6634	.6526	.6536	.6469
		0.6	.7880	.8058	.7831	—	.8022	.7910	.7927	.7884
		0.5	.8943	.8945	.8785	—	.8944	.8846	.8873	.8846

Powers of the tests for the scenario of neighbor rankings are presented in Table 2. This table shows that the test based on PA and the NP test are the best nonparametric tests. The test based on $\widehat{TPA}$ is the most powerful test for this imperfect ranking scenario, while the powers of the test based on $\widehat{TPS}$ are quite close. It is worth mentioning that in this scenario, the power difference among the proposed tests and the other tests are considerable in most cases.

Table 2: Power estimates of different level 0.05 tests, under a fraction λ of neighbor rankings under assumption of normality.

k	n	λ	$\widehat{TPS}$ (new)	$\widehat{TPA}$ (new)	$\widehat{D}$ (Z., *et al.* 2014)	MP (Frey and Wang, 2013)	NP (Frey *et al.*, 2007)	W^* (Frey *et al.*, 2007)	J (V. and Balakr., 2011)	PA (Z. *et al.*, 2012)
2	8	0.2	.1943	.1868	.1862	.1775	.1706	.1590	.1590	.1590
		0.4	.3791	.3796	.3731	.3499	.3444	.3243	.3243	.3243
		0.6	.5689	.5694	.5666	.5369	.5345	.5134	.5134	.5134
		0.8	.7328	.7358	.7350	.7066	.7062	.6897	.6897	.6897
		1	.8486	.8587	.8573	.8377	.8377	.8275	.8275	.8275
3	4	0.2	.1349	.1369	.1249	.1285	.1248	.1189	.1201	.1212
		0.4	.2483	.2524	.2164	.2371	.2266	.2122	.2173	.2206
		0.6	.3985	.4057	.3287	.3679	.3475	.3230	.3345	.3450
		0.8	.5157	.5324	.4484	.5085	.4771	.4432	.4622	.4789
		1	.6481	.6684	.5571	.6453	.6048	.5640	.5899	.6092
4	2	0.2	.1064	.1065	.0965	.0976	.0956	.0932	.0928	.0975
		0.4	.1755	.1774	.1508	.1581	.1514	.1447	.1456	.1516
		0.6	.2532	.2593	.2112	.2307	.2159	.2034	.2070	.2223
		0.8	.3416	.3514	.2774	.3136	.2875	.2678	.2757	.2987
		1	.4321	.4495	.3474	.4041	.3643	.3367	.3500	.3822
5	1	0.2	.0854	.0865	.0804	.0785	.0772	.0766	.0757	.0763
		0.4	.1275	.1282	.1156	.1109	.1066	.1051	.1040	.1064
		0.6	.1729	.1759	.1511	.1473	.1381	.1354	.1346	.1387
		0.8	.2210	.2255	.1887	.1878	.1713	.1673	.1673	.1733
		1	.2736	.2818	.2311	.2218	.2061	.2007	.2016	.2119
5	4	0.2	.1454	.1468	.1155	—	.1277	.1166	.1199	.1226
		0.4	.2733	.2799	.1983	—	.2326	.2041	.2163	.2259
		0.6	.4220	.4248	.2885	—	.3620	.3098	.3365	.3563
		0.8	.5664	.5791	.3877	—	.4959	.4266	.4668	.4954
		1	.7003	.7196	.4908	—	.6262	.5441	.5963	.6324
4	5	0.2	.1584	.1544	.1275	—	.1408	.1309	.1347	.1334
		0.4	.3072	.3051	.2260	—	.2661	.2381	.2515	.2562
		0.6	.4673	.4757	.3396	—	.4209	.3741	.4009	.4126
		0.8	.6299	.6441	.4625	—	.5750	.5123	.5537	.5716
		1	.7669	.7868	.5844	—	.7137	.6446	.6937	.7167

Partially reprinted from Computational Statistics and Data Analysis, Vol. 60, J. Frey and L. Wang, Most powerful rank tests for perfect rankings, Table 6, p. 165, Copyright 2012, with permission from Elsevier.

The simulation study was also performed for two more imperfect ranking models (fraction of inverse rankings and fraction of random rankings; see, e.g., Zamanzade *et al.*, 2014) as well as under the assumption of an exponential instead of a normal distribution. We do not report these simulation results due to space restrictions. However, they are available on request from the authors.

Remark 4.1. It is important to notice that the proposed tests have the advantage that randomization is not needed to obtain the tests of a specific size. For the nonparametric tests and the MP test, randomization is used in the simulations for a more meaningful power comparison, but when using a non-randomized version of these tests in practice, the power will be lower. For example, in the bivariate normal model, for $n = 1, k = 5$, and $\lambda = 0.5$, the estimated powers (based on 100000 repetitions) of non-randomized nonparametric tests using NP, W^*, J, and PA at a nominal level of $\alpha = 0.05$ are 0.460, 0.460, 0.435, and 0.316, respectively, which are lower that their reported values in Table 1, where the randomized tests are used.

Remark 4.2. It is worth mentioning that although the MP test has reasonably good powers in most cases, the application of this test is too restricted in practice. It should be noted that this test can only be used in practice if the underlying distribution of the population, the scenario of imperfect ranking and the fraction of imperfect rankings (λ) are all completely known. Since these conditions, especially the last one, are hardly conceivable to be satisfied, this test cannot be used in many parametrical situations in practice.

5. CONCLUSION

In this paper, we developed two parametric and location-scale free tests of perfect judgment ranking based on ordered ranked set samples. Our tests are based on the idea that if the assumption of perfect ranking is satisfied, then the difference between ranked set samples and ordered ranked set samples should be small. Then we generalized our proposed tests to the multi-cycle case of BRSS. Finally, we compared our tests with their best known competitors in the literature. Our power comparisons indicate that the proposed tests have good performance in comparison with their leading competitors, especially under the fraction of neighbor rankings model.

It is worth mentioning that although we confine ourselves to the balanced ranked set samples, the proposed tests can straightforwardly be generalized to the unbalanced case.

ACKNOWLEDGMENTS

The authors are thankful to Professor Jesse Frey for his comments on an earlier draft of this paper. The authors are also grateful to an anonymous referee and an associate editor for their valuable comments and suggestions which lead to this improved version.

REFERENCES

[1] Frey, J.; Ozturk, O. and Deshpande, J.V. (2007). Nonparametric tests for perfect judgment rankings, *Journal of the American Statistical Association*, **102**, 708–717.

[2] Frey, J. and Wang, L. (2013). Most powerful rank tests for perfect rankings, *Computational Statistics and Data Analysis*, **60**, 157–168.

[3] Li, T. and Balakrishnan, N. (2008). Some simple nonparametric methods to test for perfect ranking in ranked set sampling, *Journal of Statistical Planning and Inference*, **138**, 1325–1338.

[4] MacEachern, S.N.; Ozturk, O.; Wolfe, D.A. and Stark, G.V. (2002). A new ranked set sample estimator of variance, *Journal of the Royal Statistical Society: Series B*, **64**(2), 177–188.

[5] McIntyre, G.A. (1952). A method for unbiased selective sampling using ranked set sampling, *Australian Journal of Agricultural Research*, **3**, 385–390.

[6] McIntyre, G.A. (2005). A method for unbiased selective sampling using ranked set sampling, *The American Statistician*, **59**, 230–232. Originally appeared in Australian Journal of Agricultural Research 3, 385–390.

[7] Perron, F. and Sinha, B.K. (2004). Estimation of variance based on a ranked set sample, *Journal of Statistical Planning and Inference*, **120**(1–2), 21–28.

[8] Stokes, S.L. (1980). Estimation of variance using judgment ordered ranked set samples, *Biometrics*, **36**(1), 35–42.

[9] Vock, M. and Balakrishnan, N. (2011). A Jonckheere–Terpstra-type test for perfect ranking in balanced ranked set sampling, *Journal of Statistical Planning and Inference*, **141**, 624–630.

[10] Vock, M. and Balakrishnan, N. (2013). A connection between two nonparametric tests for perfect ranking in balanced ranked set sampling, *Communications in Statistics – Theory and Methods*, **42**, 191–193.

[11] Wolfe, D. (2012). Ranked set sampling: Its relevance and impact on statistical inference, *ISRN Probability and Statistics*, vol. 2012, Article ID 568385, 32 pages, 2012. doi:10.5402/2012/568385.

[12] Zamanzade, E.; Arghami, N.R. and Vock, M. (2012). Permutation-based tests of perfect ranking, *Statistics & Probability Letters*, **82**, 3313–2220.

[13] Zamanzade, E.; Arghami, N.R. and Vock, M. (2014). A parametric test of perfect ranking in balanced ranked set sampling, *Communications in Statistics – Theory and Methods*, **43**, 4589–4611.

REVSTAT – Statistical Journal
Volume 16, Number 4, October 2018, 475–486

HIGHLY D-EFFICIENT WEIGHING DESIGNS FOR AN EVEN NUMBER OF OBJECTS

Authors: Bronisław Ceranka
– Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland
bronicer@up.poznan.pl

Małgorzata Graczyk
– Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland
magra@up.poznan.pl

Received: April 2016 Revised: November 2016 Accepted: November 2016

Abstract:

- In this paper we formulate how to add $a = 1, 2, 3$ runs to a near D-optimal weighing design to get a highly D-efficient weighing design when the number of objects p is even.

Key-Words:

- *D-optimal design; efficiency; spring balance weighing design.*

AMS Subject Classification:

- 62K05, 05B20.

1. INTRODUCTION

We study a weighing experiment where observations follow the linear model $\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{e}$, where $\mathbf{y} = (y_1, y_2, ..., y_n)^{'}$ is a $n \times 1$ random vector of observations, $\mathbf{X}$ is the model matrix identified by the weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0, 1\}$, where $\mathbf{\Phi}_{n\times p}\{0, 1\}$ denotes the set of all $n \times p$ matrices with elements 0 or 1, $rank(\mathbf{X}) = p$, $\mathbf{w} = (w_1, w_2, ..., w_p)^{'}$ is a $p \times 1$ vector of true unknown parameters (weights) and $\mathbf{e} = (e_1, e_2, ..., e_n)^{'}$ is $n \times 1$ random vector of errors. We assume, $\mathrm{E}(\mathbf{e}) = \mathbf{0}_n$ and $\mathrm{Var}(\mathbf{e}) = \sigma^2\mathbf{I}_n$, where $\mathbf{0}_n$ is the $n \times 1$ zero vector and $\mathbf{I}_n$ is the identity matrix of order n. The least squares estimator of $\mathbf{w}$ is of the form $\hat{\mathbf{w}} = (\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}\mathbf{y}$ and the variance matrix of $\hat{\mathbf{w}}$ is given by the formula $\mathrm{Var}(\hat{\mathbf{w}}) = \sigma^2(\mathbf{X}^{'}\mathbf{X})^{-1}$ and $\mathbf{X}^{'}\mathbf{X}$ is called the information matrix for the design.
Our goal is to determine an optimal experimental plan $\mathbf{X}$ that minimizes the volume of the confidence region for $\mathbf{w}$ assuming that the errors are normally distributed. This is equivalent to the determining a design $\mathbf{X}$ such that $\det(\mathbf{X}^{'}\mathbf{X})$ is maximum. Such a design $\mathbf{X}$ is called D-optimal. D-optimality of weighing designs is studied in [3], [4], [6].

2. THE MAIN RESULT

Through the paper we assume that p is even. In [5], for even p it is shown that the maximum $\det(\mathbf{X}^{'}\mathbf{X})$ is attained if $\mathbf{X}^{'}\mathbf{X} = t\,(\mathbf{I}_p + \mathbf{J}_p)$ and each row of $\mathbf{X}$ contains k or $k + 1$ ones, where $p = 2k$ and $\mathbf{J}$ is a matrix of all 1s. For the design $\mathbf{X}$ having k ones in each row and even p, an upper bound for $\det(\mathbf{X}^{'}\mathbf{X})$ is given in [1]. In [1], the following theorem was also proven.

Theorem 2.1. *For any* $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0, 1\}$,

$$\det(\mathbf{X}^{'}\mathbf{X}) = (p-1)\left(\frac{np}{4(p-1)}\right)^p \tag{2.1}$$

if and only if

$$\mathbf{X}^{'}\mathbf{X} = \frac{n}{4(p-1)}\left(p\mathbf{I}_p + (p-2)\mathbf{J}_p\right), \tag{2.2}$$

where $\frac{np}{4(p-1)}$ *and* $\frac{n(p-2)}{4(p-1)}$ *are integers.*

Here, we define $\mathrm{D}_{\mathrm{eff}}(\mathbf{X})$ as

$$\mathrm{D}_{\mathrm{eff}}(\mathbf{X}) = \left(\frac{\det(\mathbf{X}^{'}\mathbf{X})}{\det(\mathbf{Y}^{'}\mathbf{Y})}\right)^{\frac{1}{p}}, \tag{2.3}$$

where $\mathbf{Y}$ is a regular D-optimal spring balance weighing design having k or $k+1$ ones in each row ($p = 2k$) and $\mathbf{Y}'\mathbf{Y} = \frac{(p+2)n}{4(p+1)}(\mathbf{I}_p + \mathbf{J}_p)$, see [5].

Definition 2.1. Any nonsingular spring balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ for which p is even is said to be near D-optimal if $\det(\mathbf{X}'\mathbf{X}) = (p-1)\left(\frac{np}{4(p-1)}\right)^p$.

In [1], some construction methods for near D-optimal weighing designs for certain values of n and p were provided. However, construction methods are needed for general n and p. Given a near D-optimal design for p objects and $n-a$ measurements we describe how to add a measurements in such way that the resulting design is highly D-efficient.

2.1. Adding $a = 1$ measurements

Let $\mathbf{X}_1$ be a near D-optimal design in $\mathbf{\Psi}_{(n-1)\times p}\{0,1\}$. In order to locate highly D-efficient design in $\mathbf{\Phi}_{n\times p}\{0,1\}$, we add one measurement, i.e. $p\times 1$ vector $\mathbf{x}$ of 0's or 1's having property $\mathbf{x}'\mathbf{1}_p = t$. So, $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ is given in the following form

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{x}' \end{bmatrix}. \tag{2.4}$$

Thus for $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ in (2.4), $\det(\mathbf{X}'\mathbf{X}) = \left(1 + \mathbf{x}'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}\right)\cdot\det(\mathbf{X}_1'\mathbf{X}_1)$, by Theorem 18.1.1 in [2]. Then we have the following theorem.

Theorem 2.2. *For any* $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ *given by* (2.4),

$$\det\left(\mathbf{X}'\mathbf{X}\right) \le (p-1)\left(\frac{(n-1)p}{4(p-1)}\right)^p\left(1 + \frac{p^3+8}{(n-1)p^2}\right). \tag{2.5}$$

Proof: By Theorem 2.1

$$\det(\mathbf{X}_1'\mathbf{X}_1) = (p-1)\left(\frac{(n-1)p}{4(p-1)}\right)^p \tag{2.6}$$

implies

$$\mathbf{X}_1'\mathbf{X}_1 = \frac{n-1}{4(p-1)}\left(p\mathbf{I}_p + (p-2)\mathbf{J}_p\right), \tag{2.7}$$

where $\frac{(n-1)p}{4(p-1)}$ and $\frac{(n-1)(p-2)}{4(p-1)}$ are integers. Apply the formula given in (2.6) to compute the determinant of the information matrix. So,

$$\det(\mathbf{X}'\mathbf{X}) = (p-1)\left(\frac{(n-1)p}{4(p-1)}\right)^p\left(1 + \mathbf{x}'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}\right).$$

Since $(\mathbf{X}_1'\mathbf{X}_1)^{-1} = \frac{4(p-1)}{(n-1)p}\left(\mathbf{I}_p - \frac{p-2}{p(p-1)}\mathbf{J}_p\right)$, we obtain

$$\det\left(\mathbf{X}'\mathbf{X}\right) = (p-1)\left(\frac{(n-1)p}{4(p-1)}\right)^p\left(1+\frac{4(p-1)}{(n-1)p}\left(\mathbf{x}'\mathbf{x} - \frac{p-2}{p(p-1)}\,\mathbf{x}'\mathbf{J}_p\,\mathbf{x}\right)\right). \tag{2.8}$$

To maximise (2.8), we determine the maximum value of the function

$$\eta(\mathbf{x}) = \mathbf{x}'\mathbf{x} - \frac{p-2}{p(p-1)}\mathbf{x}'\mathbf{J}_p\mathbf{x}. \tag{2.9}$$

Consequently, $\eta(\mathbf{x}) = t - \frac{p-2}{p(p-1)}t^2 \leq \frac{p^3+8}{4p(p-1)}$ and the equality holds if and only if $t = 0.5(p+2)$. From the above and (2.8) we obtain (2.5). □

Corollary 2.1. *For a spring balance weighing design* $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ *given by* (2.4), $\det\left(\mathbf{X}'\mathbf{X}\right) = (p-1)\left(\frac{(n-1)p}{4(p-1)}\right)^p\left(1+\frac{p^3+8}{(n-1)p^2}\right)$ *provided that* (2.7) *holds and* $\mathbf{x}'\mathbf{1}_p = 0.5(p+2)$.

2.2. Adding $a = 2$ measurements

Let $\mathbf{X}_1 \in \mathbf{\Phi}_{(n-2)\times p}\{0,1\}$ be near D-optimal. Let $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ be in the following form

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{x}' \\ \mathbf{y}' \end{bmatrix}, \tag{2.10}$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors of 0's and 1's and $\mathbf{x}'\mathbf{1}_p = t$, $\mathbf{y}'\mathbf{1}_p = u$, $\mathbf{x}'\mathbf{y} = m$, $0 \leq m \leq \min(t,u)$.

Theorem 2.3. *For any* $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ *given by* (2.10)

$$\det\left(\mathbf{X}'\mathbf{X}\right) \leq \begin{cases} Q(n,p)R(n,p) & \text{if } p = 0 \bmod 4 \\ Q(n,p)L(n,p) & \text{if } p+2 = 0 \bmod 4, \end{cases}$$

where

$$\begin{aligned} Q(n,p) &= (p-1)\left(\frac{(n-2)p}{4(p-1)}\right)^p, \\ R(n,p) &= \left(1+\frac{p^3+p^2+16}{(n-2)p^2}\right)\left(1+\frac{p-1}{n-2}\right), \\ L(n,p) &= \left(1+\frac{(p-1)(p+2)}{(n-2)p}\right)\left(1+\frac{(p+2)(p^2-3p+8)}{(n-2)p^2}\right). \end{aligned} \tag{2.11}$$

Proof: By Theorem 2.1

$$\det\left(\mathbf{X}_1'\mathbf{X}_1\right) = (p-1)\left(\frac{(n-2)p}{4(p-1)}\right)^p \tag{2.12}$$

implies

$$\mathbf{X}_1'\mathbf{X}_1 = \frac{n-2}{4(p-1)}\left(p\mathbf{I}_p + (p-2)\mathbf{J}_p\right), \tag{2.13}$$

where $\frac{(n-2)p}{4(p-1)}$ and $\frac{(n-2)(p-2)}{4(p-1)}$ are integers. By Theorem 18.1.1 in [2]

$$\det(\mathbf{X}'\mathbf{X}) = \det(\mathbf{X}_1'\mathbf{X}_1)\det\left(\mathbf{I}_2 + \begin{bmatrix}\mathbf{x}'\\ \mathbf{y}'\end{bmatrix}\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\begin{bmatrix}\mathbf{x} & \mathbf{y}\end{bmatrix}\right)$$

and

$$\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1} = \frac{4(p-1)}{(n-2)p}\left(\mathbf{I}_p - \frac{p-2}{p(p-1)}\mathbf{J}_p\right).$$

Next, by the formula given in (2.12) we have

$$\det(\mathbf{X}'\mathbf{X}) = (p-1)\left(\frac{(n-2)p}{4(p-1)}\right)^p \cdot \det(\mathbf{\Omega}), \tag{2.14}$$

where

$$\mathbf{\Omega} = \begin{bmatrix} 1+\frac{4(p-1)}{(n-2)p}\left(t-\frac{p-2}{p(p-1)}t^2\right) & \frac{4(p-1)}{(n-2)p}\left(m-\frac{p-2}{p(p-1)}tu\right) \\ \frac{4(p-1)}{(n-2)p}\left(m-\frac{p-2}{p(p-1)}tu\right) & 1+\frac{4(p-1)}{(n-2)p}\left(u-\frac{p-2}{p(p-1)}u^2\right)\end{bmatrix}.$$

As we want to maximise (2.14), we determine the maximum values of

$$t - \frac{p-2}{p(p-1)}t^2 \text{ and } u - \frac{p-2}{p(p-1)}u^2 \tag{2.15}$$

and concomitantly the minimum value of

$$\left(m - \frac{p-2}{p(p-1)}tu\right)^2. \tag{2.16}$$

The maximum values in (2.15) each as a function of p is attained if and only if $t = u = 0.5(p+2)$. If $p = 0 \bmod 4$, then the minimum value of (2.16) is equal to $\frac{(p^2+8)^2}{16p^2(p-1)^2}$ when $m = 0.25(p+4)$. Hence $\det(\mathbf{\Omega}) \leq \left(1+\frac{p^3+p^2+16}{(n-2)p^2}\right)\left(1+\frac{p-1}{n-2}\right)$ and

$$\det(\mathbf{X}'\mathbf{X}) \leq (p-1)\left(1+\frac{p^3+p^2+16}{(n-2)p^2}\right)\left(1+\frac{p-1}{n-2}\right)\left(\frac{(n-2)p}{4(p-1)}\right)^p. \tag{2.17}$$

The equality in (2.17) holds if and only if $t = u = 0.5(p+2)$ and $m = 0.25(p+4)$.

If $p+2 = 0 \bmod 4$, then the minimum value of (2.16) is equal to $\frac{(p+2)^2(p-4)^2}{16p^2(p-1)^2}$ when $m = 0.25(p+2)$. Therefore, $\det(\mathbf{\Omega}) \leq \left(1 + \frac{(p-1)(p+2)}{(n-2)p}\right)\left(1 + \frac{(p+2)(p^2-3p+8)}{(n-2)p^2}\right)$ and

$$\begin{aligned}\det(\mathbf{X}'\mathbf{X}) \;\leq\; & (p-1)\left(1 + \frac{(p-1)(p+2)}{(n-2)p}\right) \\ & \times \left(1 + \frac{(p+2)(p^2-3p+8)}{(n-2)p^2}\right)\left(\frac{(n-2)p}{4(p-1)}\right)^p. \end{aligned} \tag{2.18}$$

The equality in (2.18) holds if and only if $t = u = 0.5(p+2)$ and $m = 0.25(p+2)$. □

Corollary 2.2. *Let $Q(n,p)$, $R(n,p)$, $L(n,p)$ be of the form* (2.11) *and p be even. Then for a spring balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ given by* (2.10),

$$\det\left(\mathbf{X}'\mathbf{X}\right) = \begin{cases} Q(n,p)R(n,p) & \text{if } p = 0 \bmod 4 \\ Q(n,p)L(n,p) & \text{if } p+2 = 0 \bmod 4, \end{cases}$$

provided (2.13) *holds and*

$$\begin{cases} \mathbf{x}'\mathbf{1}_p = \mathbf{y}'\mathbf{1}_p = 0.5(p+2) & \\ \text{and} & \\ \mathbf{x}'\mathbf{y} = 0.25(p+4) & \text{if } \; p = 0 \bmod 4, \\ \mathbf{x}'\mathbf{y} = 0.25(p+2) & \text{if } \; p+2 = 0 \bmod 4. \end{cases}$$

2.3. Adding $a = 3$ measurements

Next, we assume that there exists a near D-optimal spring balance weighing design $\mathbf{X}_1$ for p objects and $n-3$ measurements in the class $\mathbf{\Phi}_{(n-3)\times p}\{0,1\}$. So, $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ is given in the form

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{x}' \\ \mathbf{y}' \\ \mathbf{z}' \end{bmatrix}, \tag{2.19}$$

where $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are vectors of 0's and 1's and

$$\begin{cases} \mathbf{x}'\mathbf{1}_p = t, \;\; \mathbf{x}'\mathbf{y} = m, \;\; 0 \leq m \leq \min(t,u) \\ \mathbf{y}'\mathbf{1}_p = u, \;\; \mathbf{x}'\mathbf{z} = q, \;\; 0 \leq q \leq \min(t,w) \\ \mathbf{z}'\mathbf{1}_p = w, \;\; \mathbf{y}'\mathbf{z} = h, \;\; 0 \leq h \leq \min(u,w). \end{cases} \tag{2.20}$$

By Theorem 2.1

$$\det(\mathbf{X}_1'\mathbf{X}_1) = (p-1)\left(\frac{(n-3)p}{4(p-1)}\right)^p, \tag{2.21}$$

implies

$$\mathbf{X}_1'\mathbf{X}_1 = \frac{n-3}{4(p-1)}\left(p\mathbf{I}_p + (p-2)\mathbf{J}_p\right), \tag{2.22}$$

where $\frac{n-3}{4(p-1)}$ and $\frac{(n-3)(p-2)}{4(p-1)}$ are integers. By using the formula given in (2.21) and Theorem 18.1.1 in [2], we obtain

$$\det(\mathbf{X}'\mathbf{X}) = (p-1)\left(\frac{(n-3)p}{4(p-1)}\right)^p \det\left(\mathbf{I}_3 + \begin{bmatrix}\mathbf{x}'\\ \mathbf{y}'\\ \mathbf{z}'\end{bmatrix}\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\begin{bmatrix}\mathbf{x} & \mathbf{y} & \mathbf{z}\end{bmatrix}\right).$$

Because $\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1} = \frac{4(p-1)}{(n-3)p}\left(\mathbf{I}_p - \frac{p-2}{p(p-1)}\mathbf{J}_p\right)$, we have

$$\det(\mathbf{X}'\mathbf{X}) = (p-1)\left(\frac{(n-3)p}{4(p-1)}\right)^p \det(\mathbf{T}), \tag{2.23}$$

where $\mathbf{T} = \mathbf{I}_3 + \frac{4(p-1)}{(n-3)p}\begin{bmatrix}\mathbf{x}'\\ \mathbf{y}'\\ \mathbf{z}'\end{bmatrix}\left(\mathbf{I}_p - \frac{p-2}{p(p-1)}\mathbf{J}_p\right)\begin{bmatrix}\mathbf{x} & \mathbf{y} & \mathbf{z}\end{bmatrix}$. By (2.20),

$$\begin{aligned}
\det(\mathbf{T}) \;=\; & \left(1+\frac{4(p-1)}{(n-3)p}\left(t-\frac{p-2}{p(p-1)}t^2\right)\right)\left(1+\frac{4(p-1)}{(n-3)p}\left(u-\frac{p-2}{p(p-1)}u^2\right)\right)\\
& \cdot\left(1+\frac{4(p-1)}{(n-3)p}\left(w-\frac{p-2}{p(p-1)}w^2\right)\right)\\
& +2\left(\frac{4(p-1)}{(n-3)p}\right)^3\left(m-\frac{p-2}{p(p-1)}tu\right)\left(q-\frac{p-2}{p(p-1)}tw\right)\left(h-\frac{p-2}{p(p-1)}uw\right)\\
& -\left(1+\frac{4(p-1)}{(n-3)p}\left(t-\frac{p-2}{p(p-1)}t^2\right)\right)\left(\frac{4(p-1)}{(n-3)p}\right)^2\left(h-\frac{p-2}{p(p-1)}uw\right)^2\\
& -\left(1+\frac{4(p-1)}{(n-3)p}\left(u-\frac{p-2}{p(p-1)}u^2\right)\right)\left(\frac{4(p-1)}{(n-3)p}\right)^2\left(q-\frac{p-2}{p(p-1)}tw\right)^2\\
& -\left(1+\frac{4(p-1)}{(n-3)p}\left(w-\frac{p-2}{p(p-1)}w^2\right)\right)\left(\frac{4(p-1)}{(n-3)p}\right)^2\left(m-\frac{p-2}{p(p-1)}tu\right)^2.
\end{aligned}$$

As we want to maximise (2.23), we simultaneously determine the maximum values of

$$t-\frac{p-2}{p(p-1)}t^2,\; u-\frac{p-2}{p(p-1)}u^2 \;\text{ and }\; w-\frac{p-2}{p(p-1)}w^2 \tag{2.24}$$

and the minimum values of

$$\left(h-\frac{p-2}{p(p-1)}uw\right)^2,\; \left(q-\frac{p-2}{p(p-1)}tw\right)^2 \;\text{ and }\; \left(m-\frac{p-2}{p(p-1)}tu\right)^2. \tag{2.25}$$

The maximum values in (2.24) are all attained if and only if $t = u = w = 0.5(p+2)$. If $p = 0 \bmod 4$, then the minimum values in (2.25) are equal to $\frac{(p^2+8)^2}{16p^2(p-1)^2}$ when $m = q = h = 0.25(p+4)$. Then

$$\begin{aligned}\det(\mathbf{T}) &\le \left(1+\frac{p^3+8}{(n-3)p^2}\right)^3 + 2\left(\frac{p^2+8}{(n-3)p^2}\right)^3 - 3\left(1+\frac{p^3+8}{(n-3)p^2}\right)\left(\frac{p^2+8}{(n-3)p^2}\right)^2 \\ &= \left(1-\frac{p-1}{n-3}\right)\left(\left(1+\frac{p^3+8}{(n-3)p^2}\right)\left(1+\frac{p^3+p^2+16}{(n-3)p^2}\right) - 2\left(\frac{p^2+8}{(n-3)p^2}\right)^2\right)\end{aligned}$$

and

$$\begin{aligned}\det(\mathbf{X}'\mathbf{X}) \le\ & (p-1)\left(\frac{(n-3)p}{4(p-1)}\right)^p \left(1+\frac{p-1}{n-3}\right) \\ & \cdot\left(\left(1+\frac{p^3+8}{(n-3)p^2}\right)\left(1+\frac{p^3+p^2+16}{(n-3)p^2}\right) - 2\left(\frac{p^2+8}{(n-3)p^2}\right)^2\right).\end{aligned} \tag{2.26}$$

The equality in (2.26) holds if and only if $t = u = w = 0.5(p+2)$ and $m = q = h = 0.25(p+4)$.

If $p+2 = 0 \bmod 4$, then the minimum values in (2.25) are all equal to $\frac{(p+2)^2(p-4)^2}{16p^2(p-1)^2}$ when $m = q = h = 0.25(p+2)$. An easy computation shows that

$$\begin{aligned}\det(\mathbf{T}) &\le \left(1+\tfrac{p^3+8}{(n-3)p^2}\right)^3 - 2\left(\tfrac{(p+2)(p-4)}{(n-3)p^2}\right)^3 - 3\left(1+\tfrac{p^3+8}{(n-3)p^2}\right)\left(\tfrac{(p+2)(p-4)}{(n-3)p^2}\right)^2 \\ &= \left(1+\tfrac{(p-1)(p+2)}{(n-3)p}\right)\left(\left(1+\tfrac{p^3+8}{(n-3)p^2}\right)\left(1+\tfrac{(p+2)(p^2-3p+8)}{(n-3)p^2}\right) - 2\left(\tfrac{(p+2)(p-4)}{(n-3)p^2}\right)^2\right)\end{aligned}$$

and consequently

$$\begin{aligned}\det(\mathbf{X}'\mathbf{X}) \le\ & (p-1)\left(\tfrac{(n-3)p}{4(p-1)}\right)^p \left(1+\tfrac{(p-1)(p+2)}{(n-3)p}\right) \\ & \cdot\left(\left(1+\tfrac{p^3+8}{(n-3)p^2}\right)\left(1+\tfrac{(p+2)(p^2-3p+8)}{(n-3)p^2}\right) - 2\left(\tfrac{(p+2)(p-4)}{(n-3)p^2}\right)^2\right).\end{aligned} \tag{2.27}$$

The equality in (2.27) holds if and only if $t = u = w = 0.5(p+2)$ and $m = q = h = 0.25(p+2)$. So, the following theorem is obtained.

Theorem 2.4. *For any* $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ *given by* (2.19)

$$\det\left(\mathbf{X}'\mathbf{X}\right) \le \begin{cases} W(n,p)S(n,p) & \text{if } p = 0 \bmod 4 \\ W(n,p)Q(n,p) & \text{if } p+2 = 0 \bmod 4, \end{cases} \tag{2.28}$$

where

$$\begin{aligned} W(n,p) &= (p-1)\left(\tfrac{(n-3)p}{4(p-1)}\right)^p, \\ S(n,p) &= \left(1+\tfrac{p-1}{n-3}\right)\left[\left(1+\tfrac{p^3+8}{(n-3)p^2}\right)\left(1+\tfrac{p^3+p^2+16}{(n-3)p^2}\right) - 2\left(\tfrac{p^2+8}{(n-3)p^2}\right)^2\right], \\ Q(n,p) &= \left(1+\tfrac{(p-1)(p+2)}{(n-3)p}\right)\left[\left(1+\tfrac{p^3+8}{(n-3)p^2}\right)\left(1+\tfrac{(p+2)(p^2-3p+8)}{(n-3)p^2}\right) - 2\left(\tfrac{(p+2)(p-4)}{(n-3)p^2}\right)^2\right]. \end{aligned} \tag{2.29}$$

Corollary 2.3. *Let* $W(n,p)$, $S(n,p)$, $Q(n,p)$ *be of the form* (2.29) *and* $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$ *by* (2.19). *Then*

$$\det\left(\mathbf{X}'\mathbf{X}\right) = \begin{cases} W(n,p)S(n,p) & \text{if } p = 0 \bmod 4 \\ W(n,p)Q(n,p) & \text{if } p+2 = 0 \bmod 4 \end{cases}$$

provided that (2.22) *holds and*

$$\begin{cases} \mathbf{x}'\mathbf{1}_p = \mathbf{y}'\mathbf{1}_p = \mathbf{z}'\mathbf{1}_p = 0.25(p+2) \\ \text{and} \\ \mathbf{x}'\mathbf{y} = \mathbf{x}'\mathbf{z} = \mathbf{y}'\mathbf{z} = 0.25(p+4) \qquad \text{if } p = 0 \bmod 4 \\ \mathbf{x}'\mathbf{y} = \mathbf{x}'\mathbf{z} = \mathbf{y}'\mathbf{z} = 0.25(p+2) \qquad \text{if } p+2 = 0 \bmod 4. \end{cases}$$

Some construction methods of $\mathbf{X}_1$ satisfying 2.2 are based on the incidence matrix of a balanced incomplete block design, see [1], Theorem 4. Such a matrix $\mathbf{X}_1$ exists only for certain values of p and n. Hence, if $\mathbf{X}_1$ does not exist in $\mathbf{\Phi}_{n\times p}\{0,1\}$ but exists among $\mathbf{\Phi}_{n-1\times p}\{0,1\}$, $\mathbf{\Phi}_{n-2\times p}\{0,1\}$ or $\mathbf{\Phi}_{n-3\times p}\{0,1\}$, then we can construct a highly D-efficient spring balance weighing design $\mathbf{X} \in \mathbf{\Phi}_{n\times p}\{0,1\}$. This construction is based on corollaries 2.2, 2.3 and 2.4.

3. EXAMPLES

Example 3.1. Consider the problem of weighing $p = 4$ objects in $n = 7$ measurements. Since $\frac{np}{4(p-1)} = \frac{7}{3}$ and $\frac{n(p-2)}{4(p-1)} = \frac{7}{6}$ are not integers, the matrix $\mathbf{X} \in \mathbf{\Phi}_{7\times 4}\{0,1\}$ for which (2.2) is satisfied does not exist. Now, let $\mathbf{X}_1$ be a matrix for $p = 4$ objects and $n-1 = 6$ measurements. Then $\frac{(n-1)p}{4(p-1)} = 2$, $\frac{(n-1)(p-2)}{4(p-1)} = 1$ and for

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \tag{3.1}$$

the condition (2.2) is fulfilled. By Corollary 2.1, the design $\mathbf{X} \in \mathbf{\Phi}_{7\times 4}\{0,1\}$ of the form $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ 1\ 1\ 1\ 0 \end{bmatrix}$ is highly D-efficient.

Example 3.2. By Corollary 2.2, $\mathbf{X} \in \mathbf{\Phi}_{8\times 4}\{0,1\}$ such that $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ 1\ 1\ 1\ 0 \\ 1\ 1\ 0\ 1 \end{bmatrix}$, where $\mathbf{X}_1$ is given in (3.1), is highly D-efficient for weighing 4 objects in 8 measurements.

Example 3.3. In order to weigh 4 objects in $n = 9$ measurements, let $\mathbf{X} \in \mathbf{\Phi}_{9\times 4}\{0,1\}$ be of the form $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ 1\,1\,1\,0 \\ 1\,1\,0\,1 \\ 1\,0\,1\,1 \end{bmatrix}$, where $\mathbf{X}_1$ is of the form (3.1). Hence $\mathbf{X}$ is highly D-efficient.

Example 3.4. Consider the problem of measuring 6 objects in $n = 11$ measurements. Since $\frac{np}{4(p-1)} = \frac{33}{10}$ is not an integer, the matrix $\mathbf{X} \in \mathbf{\Phi}_{11\times 6}\{0,1\}$ for which (2.2) is satisfied does not exist. Now, let $\mathbf{X}_2$ be a matrix for $p = 6$ objects and $n - 1 = 10$ measurements. In this case $\frac{(n-1)p}{4(p-1)} = 3$ and $\frac{(n-1)(p-2)}{4(p-1)} = 2$ and for the matrix

$$\mathbf{X}_2 = \begin{bmatrix} 1\,1\,0\,0\,1\,0 \\ 1\,1\,0\,0\,0\,1 \\ 1\,0\,1\,1\,0\,0 \\ 1\,0\,1\,0\,0\,1 \\ 1\,0\,0\,1\,1\,0 \\ 0\,1\,1\,1\,0\,0 \\ 0\,1\,1\,0\,1\,0 \\ 0\,1\,0\,1\,0\,1 \\ 0\,0\,1\,0\,1\,1 \\ 0\,0\,0\,1\,1\,1 \end{bmatrix} \tag{3.2}$$

the condition (2.2) is fulfilled. By Corollary 2.1, the design $\mathbf{X} \in \mathbf{\Phi}_{11\times 6}\{0,1\}$ of the form $\mathbf{X} = \begin{bmatrix} \mathbf{X}_2 \\ 1\,1\,1\,1\,0\,0 \end{bmatrix}$ is highly D-efficient.

Example 3.5. For weighing $p = 6$ objects using $n = 12$ measurements the design $\mathbf{X} \in \mathbf{\Phi}_{12\times 6}\{0,1\}$ of the form $\mathbf{X} = \begin{bmatrix} \mathbf{X}_2 \\ 1\,1\,1\,1\,0\,0 \\ 1\,1\,0\,0\,1\,1 \end{bmatrix}$ is highly D-efficient, by Corollary 2.2.

Example 3.6. For weighing $p = 6$ objects in $n = 13$ measurements $\mathbf{X} \in \mathbf{\Phi}_{13\times 4}\{0,1\}$ of the form $\mathbf{X} = \begin{bmatrix} \mathbf{X}_2 \\ 1\,1\,1\,1\,0\,0 \\ 1\,1\,0\,0\,1\,1 \\ 0\,0\,1\,1\,1\,1 \end{bmatrix}$, where $\mathbf{X}_1$ is given in (3.2), is highly D-efficient, by Corollary 2.3.

4. DISCUSSION

For each p and n, the resulting $\mathrm{D}_{\mathrm{eff}}$ based on the provided designs in Theorem 2.2, 2.3 and 2.4 are summarized in Table 1.

Table 1: $\mathrm{D}_{\mathrm{eff}}(\mathbf{X})$ of the design $\mathbf{X}$ for each p and n.

$p=4$					
n	6	7	8	9	10
$\mathrm{D}_{\mathrm{eff}}(\mathbf{X})$	0.9779	0.9641	0.9652	0.9779	1
$p=6$					
n	10	11	12	13	14
$\mathrm{D}_{\mathrm{eff}}(\mathbf{X})$	0.9927	0.9783	0.9719	0.9723	1
$p=8$					
n	14	15	16	17	18
$\mathrm{D}_{\mathrm{eff}}(\mathbf{X})$	0.9968	0.9849	0.9776	0.9701	1

ACKNOWLEDGMENTS

The authors wish to express their gratitude to an anonymous Reviewer for many helpful suggestions and comments that improved the readability of the manuscript.

REFERENCES

[1] Ceranka, B. and Graczyk, M. (2016). Recent developments in D-optimal spring balance weighing designs, *Brazilian Journal of Probability and Statistics*, to appear.

[2] Harville, D.A. (1997). *Matrix Algebra from Statistician's Perspective*, Springer-Verlag, New York, Inc.

[3] Hudelson, M.; Klee, V. and Larman, D. (1996). Largest j-simplices in d-cubes: Some relatives to the Hadamard determinant problem, *Linear Algebra and its Applications*, **24**, 519–598.

[4] Jacroux, M. and Notz, W. (1983). On the optimality of spring balance weighing designs, *The Annals of Statistics*, **11**, 970–978.

[5] Neubauer, M.G.; Watkins, W. and Zeitlin, J. (1997). Maximal j-simpplices in the real d-dimensional unit cube, *Journal of Combinatorial Theory, Ser. A*, **80**, 1–12.

[6] Neubauer, M.G.; Watkins, W. and Zeitlin, J. (1998). Notes on D-optimal designs, *Linear Algebra and its Applications*, **280**, 109–127.

REVSTAT – Statistical Journal
Volume 16, Number 4, October 2018, 487–514

ON FITTING GENERALIZED LINEAR MIXED EFFECTS MODELS FOR LONGITUDINAL BINARY DATA USING DIFFERENT CORRELATION STRUCTURES

Authors: M. Salomé Cabral
– CEAUL and Departamento de Estatística e Investigação Operacional, Faculdade de Ciências da Universidade de Lisboa, Portugal
salome@fc.ul.pt

M. Helena Gonçalves
– CEAUL and Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade do Algarve, Portugal
mhgoncal@ualg.pt

Received: January 2016 Revised: October 2016 Accepted: October 2016

Abstract:

- The generalized linear mixed effects model (GLMM) approach is widely used to analyze longitudinal binary data when the goal of the study is a subject-specific interpretation because it allows missing values on the response, provided they are missing at random (MAR), and accounts the correlation among the repeated observations of the same subject by the inclusion of random effects in the linear predictor. However, in GLMM it is assumed that the observations of the same subject are independent conditional to the random effects and covariates which may be not true. To overcome this problem [9] extended this model using binary Markov chains as the basic stochastic mechanism. The aim of this paper is to give a statistical assessment of both approaches in terms of properties such as efficiency and coverage probability, as well as, to give some guidelines for the choice of the statistical approach to an applied researcher. Both procedures are described and a simulation study is carried out to compare their performance. An analysis of a longitudinal binary data set illustrates the performance of both procedures in a practical example. The `R` packages `lme4` and `bild` are used.

Key-Words:

- *binary longitudinal data; exact likelihood; random effects; Markov chain; missing data.*

AMS Subject Classification:

- 60J10, 62J99.

1. INTRODUCTION

Longitudinal binary data studies are a powerful design and they have become increasingly popular in a wide range of applications across all disciplines. In these studies repeated observations of a response variable are taken over time on each subject in one or more treatment groups. In such cases the repeated measures of each vector of responses are likely to be correlated and the autocorrelation structure for the repeated data plays a significant role in the estimation of regression parameters. Although longitudinal studies are design to collect data on every subject in the sample at each time of follow-up, many studies have missing data since it is difficult to have complete records of all subjects for a variety of reasons. When longitudinal binary data are incomplete there are important implications for their analysis and one of the main concerns is to distinguish different reasons of missingness. The nature of missing data mechanism has been classified by [16] and [13] as: missing completely at random (MCAR), missing at random (MAR) and non missing at random (NMAR). Another important distinction is whether missing values occur intermittently or as dropouts. When missing values occur as dropouts, an individual is observed only at a certain time and misses all the subsequent observations. When missing values occur as intermittently, an individual may miss some measurement times among a common set of predefined measurement times. To all these situations several methods have been proposed ([4], [5], [1], [6], [14]). A review of this topic is given in [12].

In [12] is argued that methods based on likelihood, such generalized linear mixed effects model [3], usually denoted by GLMM, are recommended when the goal of the study is a subject-specific interpretation and missing values are allowed on the response, provide they are MAR in the standard terminology of [16]. In the GLMM the correlation among the repeated observations of the same subject is account by the inclusion of random effects in the linear predictor and it is assumed that observations to the same subject are independent conditional to the random effects and covariates. Although in GLMM this independence is assumed they may still be correlated. To overcome this problem [9] used a binary Markov chain model to accommodate serial dependence and odds-ratio to measure dependence between successive observations. This methodology is a development of the alternative likelihood-based formulation for a logistic regression presented by [1] which allows: (i) a first order and a second order Markov dependence; (ii) a random intercept term in the linear predictor; (iii) missing values on the response, provided they are MAR. Both approaches, GLMM and generalized linear mixed effects model with binary Markov chain (GLM3C) as the basic stochastic mechanism, are implemented in `R` [18] packages. The goal of this paper is to give information to the practitioners about which of the two procedures, GLMM or GLM3C, is more appropriate to use for their data at hand.

To achieved that goal a simulation study was carried out to compare the two aforementioned approaches in terms of properties such as efficiency and coverage probability. For GLM3C approach the estimates were obtained through the `bild` function of the `R` package `bild`. When the GLMM approach was used the estimates were achieved through the `glmer` function of the `R` package `lme4` [2] as well as the `bild` function of `R` package `bild` [11] with the independence structure selected.

The paper is organized as follows: Section 2 gives a summary of the models used. Section 3 reports a simulation study to assess the performance of the procedures. In Section 4 a real data is used to illustrate the two procedures as well as the key results of the simulation study. Section 5 concludes the paper.

2. MODEL FOR BINARY DATA

Suppose that n independent individuals are observed at times $t = 1, ..., T_i$, which need not be the same for all n individual and, to establish notation, denote by $y_{it} \in \{0, 1\}$ the binary response value at time t from individual i $(i = 1, ..., n)$, and by Y_{it} its generating random variable whose mean value is $\Pr(Y_{it} = 1) = \theta_{it}$. The sequence $(y_{i1}, ..., y_{iT_i})$ will be collectively referred as the i-th individual profile and associated with each observation time and each subject, a set of p covariates is available, denoted by x_{it}.

The logistic regression model which links the covariates and the marginal mean of Y_{it} assumes the form

$$\operatorname{logit} \theta_{it} = x_{it}^{\top} \beta, \tag{2.1}$$

where β is the p-dimensional parameter of interest and $\operatorname{logit} \theta = \log\{\theta/(1-\theta)\}$.

In longitudinal studies the repeated measures of each vector of responses are likely to be correlated. To account for the within-subject association the GLMM uses random effects, b_i, in the linear predictor. The correlation among observations from one subject can be thought of as arising from sharing a set of underlying random effects.

In what follows only the random intercept model is considered.

2.1. Generalized linear mixed effects model

The introduction of random effects can be formulated by adding a $q \times 1$ vector b_i of random effects in (2.1) associated to a $q \times 1$ vector of covariates, z_{it}, (in general a subset of $\mathbf{x}_{it}$). In the random intercept model the vector b_i is reduced to a single ($q = 1$) random effect $b_i \sim N(0, \sigma^2)$ and $z_{it} = 1$ for all $i = 1, ..., n$ and $t = 1, ..., T_i$ leading to

$$\text{logit}\,\Pr(Y_{it} = 1|b_i) = x_{it}^\top \beta + b_i, \qquad (i = 1, ..., n) \tag{2.2}$$

where the b_i's are assumed to be sampled independently from each other and that conditioning on x_{it} and b_i, the Y_{it}'s are independent.

The likelihood inference is based on a sample of n individual profiles that are assumed to be independent from each other. The contribution of the i-th subject to the likelihood of the random intercept model is

$$L_i^R(\beta, \omega) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{\mathbb{R}} L_i^F(\beta^{b_i}|b_i) \exp\left(-\frac{b_i^2}{2\,\sigma^2}\right) \mathrm{d}b_i \tag{2.3}$$

where β^{b_i} is a p-vector of parameters like β, but where the first component is now $\beta_0 + b_i$, instead of β_0 and $\omega = \log \sigma^2$. In expression (2.3) the term $L_i^F(\beta^{b_i}|b_i) = \exp\{\ell_i^F(\beta^{b_i}|b_i)\}$ where

$$\ell_i^F(\beta^{b_i}|b_i) = \sum_{t=1}^{T_i} \big[y_{it}\,\text{logit}(\theta_{it}) + \log(1 - \theta_{it})\big].$$

The log-likelihood for the whole sample is given by

$$\ell^R(\beta, \omega) = \sum_{i=1}^{n} \log L_i^R(\beta, \omega). \tag{2.4}$$

The integrals in (2.3) have no analytical solution and appropriate numerical integration methods must be used.

This methodology is implemented in the `R` package `bild` [11] and the integrals in (2.3) are computed using adaptive Gaussian quadrature. Other `R` packages have this procedure implemented and one of the most popular is the `lme4` [2] package that also uses adaptive Gaussian quadrature to compute the integrals in (2.3).

2.2. Generalized linear mixed effects model with Markov chain correlation

Although in GLMM it is assumed that conditioning on x_{it} and b_i, the Y_{it}'s are independent they may still be correlated. To overcome this problem [9] proposed the use of binary by Markov chains to model the serial dependence between successive observations of the same subject. As they note Markov chains provide the simplest stochastic mechanism to introduce serial dependence for discrete random variables. In their approach the serial dependence between successive observation can be regulated (i) by one dependence parameter (first order dependence structure) or (ii) by two dependence parameters the (second order dependence structure). In both cases the odds-ratio is the quantity used to measure dependence between variables. One advantage of odds-ratios as measures of association is that, unlike marginal correlations, they are not constrained by marginal probabilities ([1], [8]). Their approach can be summarized as follows. To simplify notation the subscript i is dropped temporarily.

For the first order dependence structure (MC1), the serial dependence is modeled using $\psi_1 = OR(Y_t, Y_{t-1})$ where

$$OR(Y_t, Y_{t-1}) = \frac{\Pr(Y_{t-1} = Y_t = 1)\ \Pr(Y_{t-1} = Y_t = 0)}{\Pr(Y_{t-1} = 0,\ Y_t = 1)\ \Pr(Y_{t-1} = 1,\ Y_t = 0)} = \frac{p_1/(1-p_1)}{p_0/(1-p_0)}$$

where p_j are the transition probabilities given by

$$p_j = \Pr(Y_t = 1 | Y_{t-1} = j), \quad j = 0, 1; t = 2, ..., T. \tag{2.5}$$

For the second order dependence structure (MC2) is considered the joint distribution of three components of the process at time, (Y_{t-2}, Y_{t-1}, Y_t) and impose the constraints

$$\begin{aligned} OR(Y_{t-1}, Y_{t-2}) &= \psi_1 = OR(Y_t, Y_{t-1}) \\ OR(Y_t, Y_{t-2} | Y_{t-1} = 0) &= \psi_2 = OR(Y_t, Y_{t-2} | Y_{t-1} = 1) \end{aligned}$$

ψ_1 and ψ_2 denote two positive parameters. The transition probabilities are given by

$$p_{hj} = \Pr(Y_t = 1 | Y_{t-2} = h, Y_{t-1} = j), \quad h, j = 0, 1; t = 3, ..., T, \tag{2.6}$$

see [8] for a full account.

The serial dependence for MC2 models is regulated by $\lambda = (\lambda_1, \lambda_2) = (\log \psi_1, \log \psi_2)$, which are assume to be constant across time and subjects. When $\lambda_2 = 0$, the Markov chain reduces to MC1 models and the serial dependence is regulated by λ_1.

The likelihood inference is based on a sample of n individual profiles that are assumed to be independent from each other. The contribution of the i-th subject to the likelihood of the random intercept model is

$$L_i^R(\beta, \lambda, \omega) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{\mathbb{R}} L_i^F(\beta^{b_i}, \lambda | b_i) \exp\left(-\frac{b_i^2}{2\,\sigma^2}\right) \mathrm{d}b_i \tag{2.7}$$

where β^{b_i} and ω are defined as in Section 2.1. In expression (2.7) the term $L_i^F(\beta^{b_i}, \lambda | b_i) = \exp\{\ell_i^F(\beta^{b_i}, \lambda | b_i)\}$ is computed, under a serial dependence MC1, from

$$\ell_i^F(\beta, \lambda) = y_1 \operatorname{logit}(\theta_1) + \log(1 - \theta_1) + \sum_{t=2}^{T_i} [y_t \operatorname{logit}(p_j) + \log(1 - p_j)] \tag{2.8}$$

and under a serial dependence MC2 from

$$\begin{aligned} \ell_i^F(\beta, \lambda) \;=\; & \big[y_1 \operatorname{logit}(\theta_1) + \log(1 - \theta_1)\big] + \big[y_2 \operatorname{logit}(p_j) + \log(1 - p_j)\big] \\ & + \sum_{t=3}^{T_i} \Big[y_t \operatorname{logit}(p_{hj}) + \log(1 - p_{hj})\Big] \end{aligned} \tag{2.9}$$

where the three blocks on the right-hand side represent the contribution to the log-likelihood from y_1, y_2, and $(y_3, ..., y_T)$, respectively, where p_{hj} is given by (2.6) and p_j by (2.5). The log-likelihood for the whole sample is given by (2.4). For a full account see [8].

In this approach missing values are allowed on the response, provided they are MAR. If missing data occur at the beginning or at the end of an individual profile, this poses no problems, since this case is equivalent to a designed unbalance in the length profile T_i for that individual. Some restrictions exist for the presence of missing data when they occur in the middle of the profile due to the imposed correlation structure. If MC1 model is considered and if there is a missing value at time point $t - 1$, it is required that there are observations at time points $t - 2$ and t. If MC2 model is considered and if there is a missing value at time point $t - 2$, it is required that there are observations at time points $t - 4, t - 3, t - 1$ and t, except for the two end portions of the observation period, where no restriction is made.

This approach is implemented in the `R` package `bild` ([11]) and, as in the previous approach, the integrals in (2.7) are computed using adaptive Gaussian quadrature.

3. A SIMULATION STUDY

A simulation study was carried out to compare both approaches when simulated data has a serial dependence MC1 or MC2. The model considered in the simulation included a dichotomous treatment, a linear effect time and an interaction between time and treatment and is given by

$$\Pr(Y_{it}=1|t)=\frac{\exp(\beta_0+b_i+\beta_1 t+\beta_2 x_i+\beta_3(x_i\times t))}{1+\exp(\beta_0+b_i+\beta_1 t+\beta_2 x_i+\beta_3(x_i\times t))} \tag{3.1}$$

where $x_i=0$ for half the population and 1 for the remainder. The fixed effect coefficients were set at $\beta_0=-1$, $\beta_1=0.5$, $\beta_2=1$, $\beta_3=1$ and the random effect distribution was simulated with $b_i\sim N(0,\sigma^2)$. In both serial dependence structures several designs were considered to reflect the range of experimental data encountered in practice. The number of subjects was set to either small ($n=20$) or large ($n=50$). The length of profile on each subject was short ($T=7$) or long ($T=13$) and the time points were set for $T=7$ at $t=-1.5,-1,-0.5,0,0.5,1,1.5$ and for $T=13$ at $t=-1.5,-1.25,-1,-0.75,-0.5,-0.25,0,0.25,0.5,0.75,1,1.25,1.5$.

In what concerns the variance of the random effect, σ^2, three values were considered $\sigma^2=0.5,1$ and 2.

1. Under MC1 models on each run were generated T binary correlated data under the i-th subject following a first order serial dependence regulated by λ_1. The values considered for λ_1 were 0.05, 0.25, 0.5, 0.75 and 1.
2. Under MC2 models on each run were generated T binary correlated data under the i-th subject following a second order serial dependence regulated by $\lambda=(\lambda_1,\lambda_2)$. For the pair (λ_1,λ_2) the combinations $(0.05,0.05)$, $(0.25,0.25)$, $(0.5,0.5)$, $(0.75,0.75)$ and $(1,1)$ were considered.

In both cases the whole estimation procedure was repeated for 1000 runs and the sample mean of estimate parameter (Mean), the sample mean of bias (Bias) and the sample mean square error (MSE) were computed, as well as, the coverage probabilities of nominal 95% confidence intervals (CI). For each data set the relative efficiency (RE) of the estimators was computed, as usual, by the ratio of the respective MSE. RE>1 means GLM3C estimator is preferred. The coverage probabilities of nominal 95% confidence intervals were computed as the proportion of simulated intervals that cover the true parameter used to generate the simulated data.

For the GLM3C approach the estimates of the parameters were obtained through the function `bild` in the `R` package `bild` and the dependence structure

was chosen trough the argument `dependence` in the function `bild`, `MC1R` (MC1 with random intercept) `MC2R` (MC2 with random intercept), for details see [10]. When GLMM approach (which ignores the conditional dependence between repeated measures in terms of numerical analysis) was considered the estimates were obtained through the `bild` function with the dependence argument set at `indR` (independence structure with random intercept) as well as through the `glmer` function in the `R` package `lme4`, the results obtained were exactly the same.

The results of simulation are displayed from Figures 1–6 and Tables 1–8 for the time effect (β_1) and group-time interaction effect (β_3) the effects usually of most interest in a longitudinal study. Each table lists the following: Mean, Bias, MSE and coverage probability of nominal 95% confidence intervals for β_1 and β_3 over the 1000 simulations to both approaches (GLMM and GLMC3). The GLM3C approach is denoted by GLM3C-MC1 or GLM3C-MC2 if a serial dependence MC1 or MC2, respectively, is considered. The Figures display the results concerned to $\sigma^2 = 0.5$ and $\sigma^2 = 2$, the two extreme values considered to the variance of the random effect b_i.

Taking into account that the goal of the simulation study is to give a statistical assessment of both approaches the main conclusions to serial dependence MC1 and MC2 are given, respectively, in Sections 3.1 and 3.2.

3.1. Serial dependence MC1

Under a serial dependence MC1 and to β_1 and β_3 parameters the main conclusions, based on the approaches GLM3C-MC1 and GLMM, are:

(**i**) The coverage probabilities of both approaches are similar when the dependence structure established by λ_1 is low (0.05-0.25). When λ_1 increases the GLM3C approach gives coverage probabilities closer to nominal than the GLMM approach (Figures 1–2 and Tables 1–4).

(**ii**) When λ_1 is low (0.05) the efficiency of both approaches is similar with the GLMM estimators more efficient in some configurations. As λ_1 increases the GLM3C estimators becomes more efficient than the GLMM for all the design configurations (Figure 3 and Tables 1–4).

(**iii**) In terms of bias the behavior of both approaches is very similar with a slight decrease of the estimated bias associated with the GLM3C approach when λ_1 increases. The exception is for $\hat{\beta}_1$ when $T = 7$, $n = 50$ and for all σ^2 considered (Tables 1–4).

Table 1: Results of the simulation study under a serial dependence MC1 for $n = 20$, $T = 7$.

		GLM3C-MC1					**GLMM**				
		λ_1					λ_1				
		.05	.25	.50	.75	1	.05	.25	.50	.75	1
		$\sigma^2 = 0.5$									
Mean	$\hat{\beta}_1$	0.521	0.520	0.513	0.514	0.513	0.519	0.523	0.525	0.532	0.537
	$\hat{\beta}_3$	1.110	1.108	1.111	1.109	1.082	1.111	1.120	1.137	1.152	1.132
Bias	$\hat{\beta}_1$	0.021	0.020	0.013	0.014	0.013	0.019	0.023	0.025	0.032	0.037
	$\hat{\beta}_3$	0.110	0.108	0.111	0.109	0.082	0.111	0.120	0.137	0.152	0.132
MSE	$\hat{\beta}_1$	0.090	0.094	0.098	0.106	0.112	0.087	0.094	0.100	0.112	0.125
	$\hat{\beta}_3$	0.279	0.289	0.311	0.322	0.325	0.279	0.297	0.330	0.356	0.361
CI	β_1	0.947	0.951	0.947	0.951	0.951	0.950	0.945	0.927	0.928	0.927
	β_3	0.953	0.959	0.952	0.952	0.953	0.957	0.950	0.932	0.934	0.929
		$\sigma^2 = 1$									
Mean	$\hat{\beta}_1$	0.511	0.522	0.514	0.512	0.509	0.508	0.524	0.527	0.536	0.536
	$\hat{\beta}_3$	1.135	1.074	1.109	1.106	1.048	1.135	1.084	1.141	1.158	1.108
Bias	$\hat{\beta}_1$	0.011	0.022	0.014	0.012	0.009	0.008	0.024	0.027	0.036	0.036
	$\hat{\beta}_3$	0.135	0.074	0.109	0.106	0.048	0.135	0.084	0.141	0.158	0.108
MSE	$\hat{\beta}_1$	0.105	0.100	0.106	0.105	0.131	0.104	0.100	0.111	0.116	0.147
	$\hat{\beta}_3$	0.315	0.281	0.329	0.324	0.327	0.313	0.284	0.354	0.369	0.363
CI	β_1	0.947	0.948	0.950	0.949	0.955	0.945	0.942	0.927	0.928	0.916
	β_3	0.947	0.951	0.951	0.957	0.950	0.941	0.940	0.930	0.938	0.922
		$\sigma^2 = 2$									
Mean	$\hat{\beta}_1$	0.529	0.527	0.528	0.497	0.500	0.527	0.531	0.539	0.516	0.531
	$\hat{\beta}_3$	1.056	1.060	1.036	1.066	1.045	1.053	1.070	1.064	1.113	1.111
Bias	$\hat{\beta}_1$	0.029	0.027	0.028	−0.003	0.000	0.027	0.031	0.039	0.016	0.031
	$\hat{\beta}_3$	0.055	0.060	0.036	0.066	0.04	0.053	0.070	0.064	0.113	0.111
MSE	$\hat{\beta}_1$	0.119	0.121	0.130	0.115	0.131	0.118	0.121	0.136	0.125	0.149
	$\hat{\beta}_3$	0.308	0.292	0.326	0.304	0.315	0.300	0.294	0.347	0.334	0.3623
CI	β_1	0.955	0.948	0.955	0.944	0.948	0.953	0.949	0.949	0.914	0.897
	β_3	0.948	0.955	0.960	0.960	0.962	0.953	0.960	0.951	0.933	0.927

Table 2: Results of the simulation study under a serial dependence MC1 for $n = 20$, $T = 13$.

		GLM3C-MC1					**GLMM**				
		λ_1					λ_1				
		.05	.25	.50	.75	1	.05	.25	.50	.75	1
		$\sigma^2 = 0.5$									
Mean	$\hat{\beta}_1$	0.500	0.502	0.497	0.495	0.496	0.501	0.505	0.505	0.508	0.516
	$\hat{\beta}_3$	1.041	1.041	1.045	1.052	1.051	1.041	1.048	1.060	1.078	1.091
Bias	$\hat{\beta}_1$	0.000	0.002	−0.003	−0.005	−0.004	0.001	0.005	0.005	0.008	0.016
	$\hat{\beta}_3$	0.041	0.041	0.045	0.052	0.051	0.041	0.048	0.060	0.078	0.091
MSE	$\hat{\beta}_1$	0.050	0.055	0.054	0.060	0.061	0.050	0.055	0.056	0.063	0.067
	$\hat{\beta}_3$	0.134	0.142	0.147	0.160	0.178	0.133	0.143	0.152	0.173	0.200
CI	β_1	0.958	0.946	0.953	0.947	0.950	0.951	0.932	0.936	0.913	0.916
	β_3	0.960	0.950	0.956	0.957	0.956	0.953	0.944	0.933	0.937	0.910
		$\sigma^2 = 1$									
Mean	$\hat{\beta}_1$	0.501	0.502	0.500	0.501	0.498	0.501	0.505	0.508	0.515	0.518
	$\hat{\beta}_3$	1.043	1.046	1.044	1.051	1.047	1.044	1.053	1.061	1.077	1.088
Bias	$\hat{\beta}_1$	0.001	0.002	0.000	0.001	−0.002	0.001	0.005	0.008	0.015	0.018
	$\hat{\beta}_3$	0.043	0.046	0.044	0.051	0.047	0.044	0.053	0.061	0.077	0.088
MSE	$\hat{\beta}_1$	0.052	0.056	0.057	0.061	0.071	0.052	0.056	0.059	0.065	0.079
	$\hat{\beta}_3$	0.142	0.150	0.158	0.166	0.184	0.142	0.153	0.166	0.181	0.207
CI	β_1	0.951	0.954	0.962	0.964	0.942	0.944	0.939	0.934	0.924	0.894
	β_3	0.964	0.960	0.960	0.968	0.957	0.959	0.949	0.938	0.934	0.913
		$\sigma^2 = 2$									
Mean	$\hat{\beta}_1$	0.510	0.504	0.506	0.506	0.508	0.510	0.507	0.513	0.519	0.528
	$\hat{\beta}_3$	1.039	1.044	1.048	1.049	1.042	1.038	1.051	1.064	1.075	1.082
Bias	$\hat{\beta}_1$	0.010	0.004	0.006	0.006	0.008	0.010	0.007	0.013	0.019	0.028
	$\hat{\beta}_3$	0.039	0.044	0.048	0.049	0.042	0.038	0.051	0.064	0.795	0.082
MSE	$\hat{\beta}_1$	0.058	0.059	0.064	0.072	0.080	0.058	0.060	0.066	0.076	0.089
	$\hat{\beta}_3$	0.175	0.180	0.194	0.213	0.225	0.173	0.183	0.202	0.229	0.255
CI	β_1	0.960	0.955	0.952	0.947	0.953	0.962	0.946	0.946	0.923	0.905
	β_3	0.954	0.957	0.946	0.946	0.945	0.960	0.956	0.939	0.921	0.907

Table 3: Results of the simulation study under a serial dependence MC1 for $n = 50$, $T = 7$.

		GLM3C-MC1					**GLMM**				
		λ_1					λ_1				
		.05	.25	.50	.75	1	.05	.25	.50	.75	1
		$\sigma^2 = 0.5$									
Mean	$\hat{\beta}_1$	0.492	0.489	0.484	0.483	0.481	0.493	0.495	0.498	0.505	0.513
	$\hat{\beta}_3$	1.068	1.062	1.056	1.048	1.040	1.070	1.074	1.084	1.092	1.102
Bias	$\hat{\beta}_1$	−0.008	−0.011	−0.016	−0.017	−0.019	−0.007	−0.005	−0.002	0.005	0.013
	$\hat{\beta}_3$	0.068	0.062	0.056	0.048	0.040	0.070	0.074	0.084	0.092	0.102
MSE	$\hat{\beta}_1$	0.032	0.033	0.036	0.037	0.039	0.032	0.033	0.038	0.040	0.045
	$\hat{\beta}_3$	0.100	0.102	0.107	0.116	0.115	0.099	0.105	0.115	0.130	0.139
CI	β_1	0.959	0.962	0.960	0.957	0.956	0.956	0.954	0.945	0.938	0.926
	β_3	0.951	0.951	0.951	0.947	0.958	0.944	0.936	0.931	0.916	0.911
		$\sigma^2 = 1$									
Mean	$\hat{\beta}_1$	0.489	0.495	0.482	0.480	0.477	0.499	0.502	0.496	0.503	0.506
	$\hat{\beta}_3$	1.057	1.045	1.062	1.055	1.021	1.058	1.057	1.091	1.100	1.088
Bias	$\hat{\beta}_1$	−0.011	−0.005	−0.018	−0.020	−0.023	−0.001	0.002	−0.004	0.003	0.006
	$\hat{\beta}_3$	0.057	0.045	0.062	0.055	0.021	0.058	0.057	0.091	0.100	0.088
MSE	$\hat{\beta}_1$	0.032	0.038	0.039	0.040	0.046	0.032	0.039	0.041	0.043	0.052
	$\hat{\beta}_3$	0.103	0.106	0.101	0.106	0.115	0.101	0.109	0.111	0.121	0.136
CI	β_1	0.953	0.948	0.947	0.957	0.955	0.949	0.938	0.938	0.937	0.916
	β_3	0.937	0.941	0.956	0.954	0.950	0.935	0.929	0.939	0.925	0.922
		$\sigma^2 = 2$									
Mean	$\hat{\beta}_1$	0.496	0.490	0.488	0.486	0.469	0.499	0.497	0.503	0.506	0.499
	$\hat{\beta}_3$	1.028	1.081	1.074	1.026	1.041	1.030	1.093	1.102	1.072	1.108
Bias	$\hat{\beta}_1$	−0.004	−0.010	−0.012	−0.014	−0.031	−0.001	−0.003	0.003	0.006	−0.001
	$\hat{\beta}_3$	0.028	0.081	0.074	0.026	0.041	0.030	0.093	0.102	0.072	0.108
MSE	$\hat{\beta}_1$	0.038	0.040	0.044	0.044	0.051	0.037	0.042	0.046	0.047	0.057
	$\hat{\beta}_3$	0.097	0.115	0.119	0.110	0.124	0.095	0.118	0.129	0.124	0.151
CI	β_1	0.948	0.962	0.953	0.945	0.942	0.938	0.955	0.929	0.926	0.893
	β_3	0.953	0.955	0.952	0.953	0.945	0.946	0.943	0.933	0.927	0.905

Table 4: Results of the simulation study under a serial dependence MC1 for $n = 50$, $T = 13$.

		GLM3C-MC1					**GLMM**				
		λ_1					λ_1				
		.05	.25	.50	.75	1	.05	.25	.50	.75	1
		$\sigma^2 = 0.5$									
Mean	$\hat{\beta}_1$	0.497	0.496	0.495	0.494	0.500	0.498	0.500	0.503	0.508	0.521
	$\hat{\beta}_3$	1.043	1.040	1.037	1.035	1.028	1.044	1.048	1.055	1.063	1.067
Bias	$\hat{\beta}_1$	−0.003	−0.004	−0.005	−0.006	0.000	−0.002	0.000	0.003	0.008	0.021
	$\hat{\beta}_3$	0.043	0.040	0.037	0.035	0.028	0.044	0.048	0.055	0.063	0.067
MSE	$\hat{\beta}_1$	0.022	0.023	0.026	0.028	0.029	0.022	0.024	0.027	0.030	0.032
	$\hat{\beta}_3$	0.057	0.060	0.064	0.070	0.074	0.058	0.062	0.068	0.078	0.084
CI	β_1	0.953	0.949	0.945	0.942	0.949	0.949	0.939	0.919	0.905	0.890
	β_3	0.942	0.943	0.943	0.944	0.951	0.937	0.932	0.926	0.910	0.900
		$\sigma^2 = 1$									
Mean	$\hat{\beta}_1$	0.496	0.496	0.496	0.495	0.493	0.497	0.500	0.504	0.508	0.513
	$\hat{\beta}_3$	1.053	1.050	1.046	1.044	1.040	1.055	1.058	1.064	1.072	1.079
Bias	$\hat{\beta}_1$	−0.004	−0.005	−0.004	−0.005	−0.007	−0.003	0.000	0.004	0.008	0.013
	$\hat{\beta}_3$	0.053	0.050	0.046	0.044	0.040	0.055	0.058	0.064	0.072	0.079
MSE	$\hat{\beta}_1$	0.022	0.025	0.028	0.029	0.030	0.022	0.025	0.029	0.031	0.033
	$\hat{\beta}_3$	0.060	0.064	0.069	0.073	0.078	0.060	0.066	0.074	0.081	0.090
CI	β_1	0.955	0.949	0.949	0.951	0.952	0.951	0.934	0.925	0.914	0.891
	β_3	0.955	0.940	0.943	0.950	0.946	0.948	0.930	0.924	0.912	0.902
		$\sigma^2 = 2$									
Mean	$\hat{\beta}_1$	0.497	0.496	0.495	0.492	0.494	0.498	0.500	0.504	0.505	0.513
	$\hat{\beta}_3$	1.059	1.058	1.062	1.060	1.055	1.061	1.066	1.080	1.090	1.096
Bias	$\hat{\beta}_1$	−0.003	−0.004	−0.005	−0.008	−0.006	−0.002	0.000	0.004	0.005	0.013
	$\hat{\beta}_3$	0.059	0.058	0.062	0.060	0.055	0.061	0.066	0.080	0.090	0.096
MSE	$\hat{\beta}_1$	0.025	0.027	0.029	0.032	0.035	0.025	0.027	0.030	0.035	0.039
	$\hat{\beta}_3$	0.068	0.070	0.077	0.085	0.088	0.066	0.072	0.082	0.095	0.103
CI	β_1	0.959	0.964	0.954	0.953	0.955	0.950	0.949	0.938	0.927	0.895
	β_3	0.955	0.952	0.953	0.944	0.950	0.951	0.940	0.928	0.909	0.896

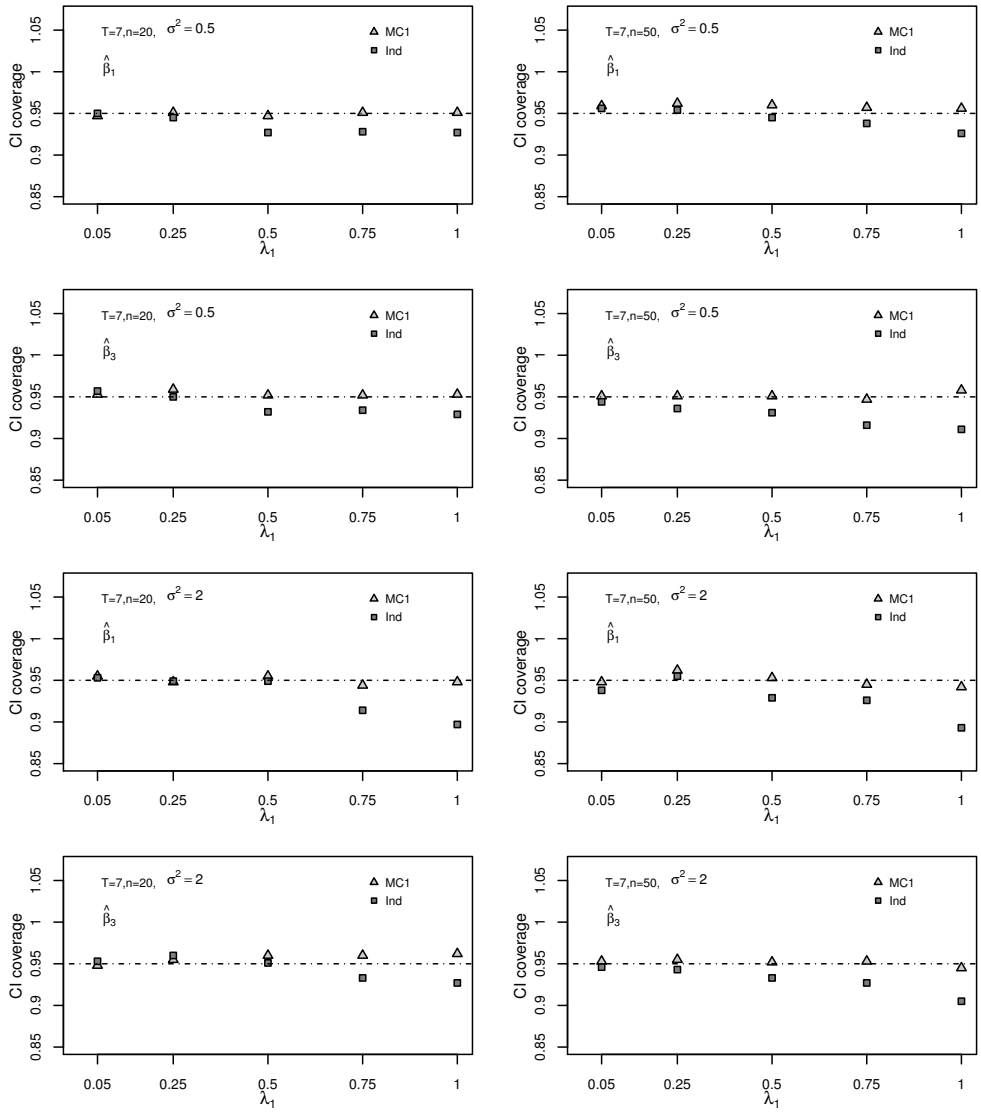


Figure 1: Coverage probabilities of nominal 95% confidence intervals (CI coverage) for β_1 ($\hat{\beta}_1$) and β_3 ($\hat{\beta}_3$) based on 1000 runs using GLM3C and GLMM estimation procedures (T = length of profile on each subject, n = number of subjects, σ^2 = variance of the random effect). Data set on each run has a serial dependence MC1 regulated by λ_1. Coding for estimation procedures: MC1 (GLM3C) and Ind (GLMM).

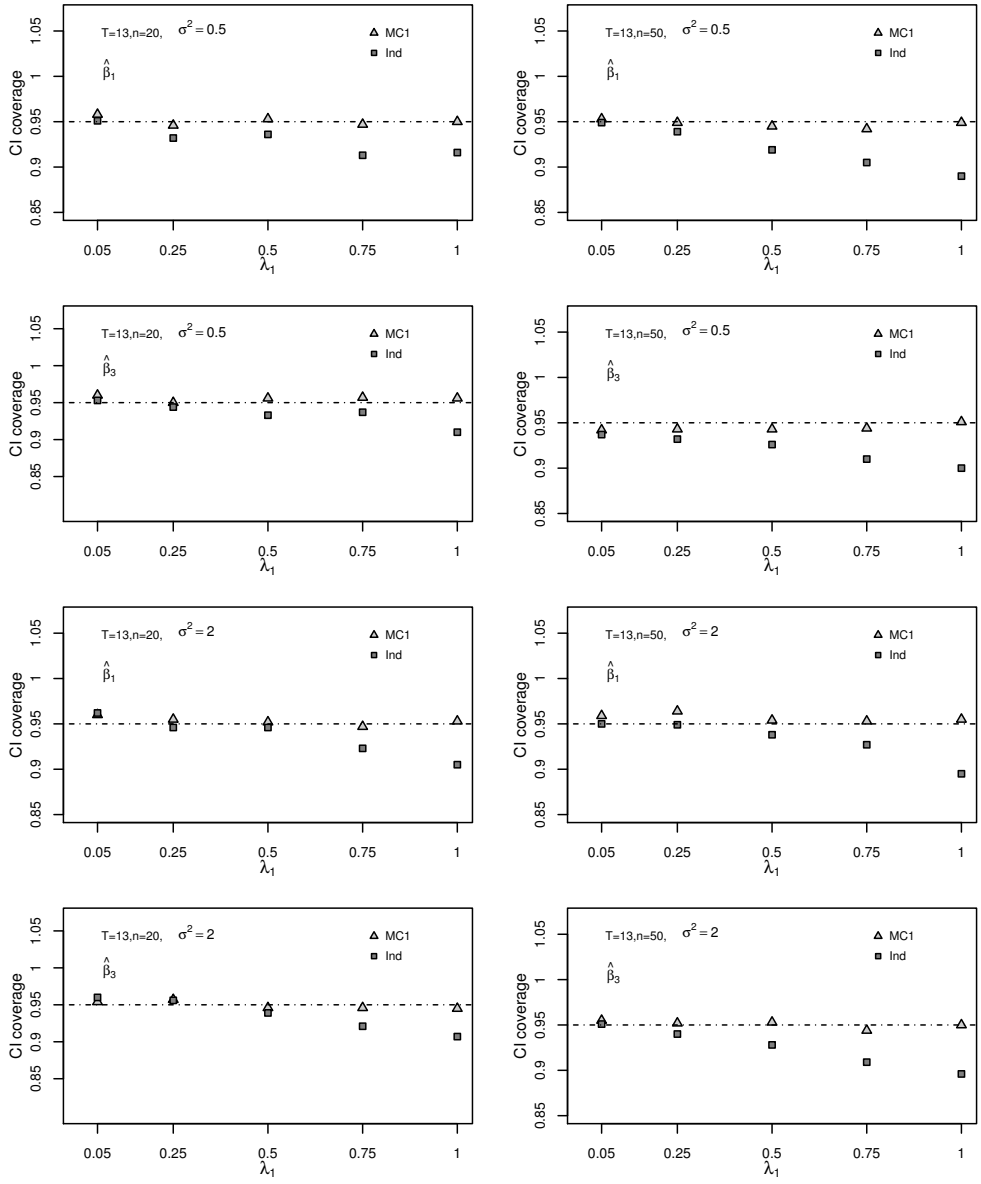


Figure 2: Coverage probabilities of nominal 95% confidence intervals (CI coverage) for β_1 ($\hat{\beta}_1$) and β_3 ($\hat{\beta}_3$) based on 1000 runs using GLM3C and GLMM estimation procedures (T = length of profile on each subject, n = number of subjects, σ^2 = variance of the random effect). Data set on each run has a serial dependence MC1 regulated by λ_1. Coding for estimation procedures: MC1 (GLM3C) and Ind (GLMM).

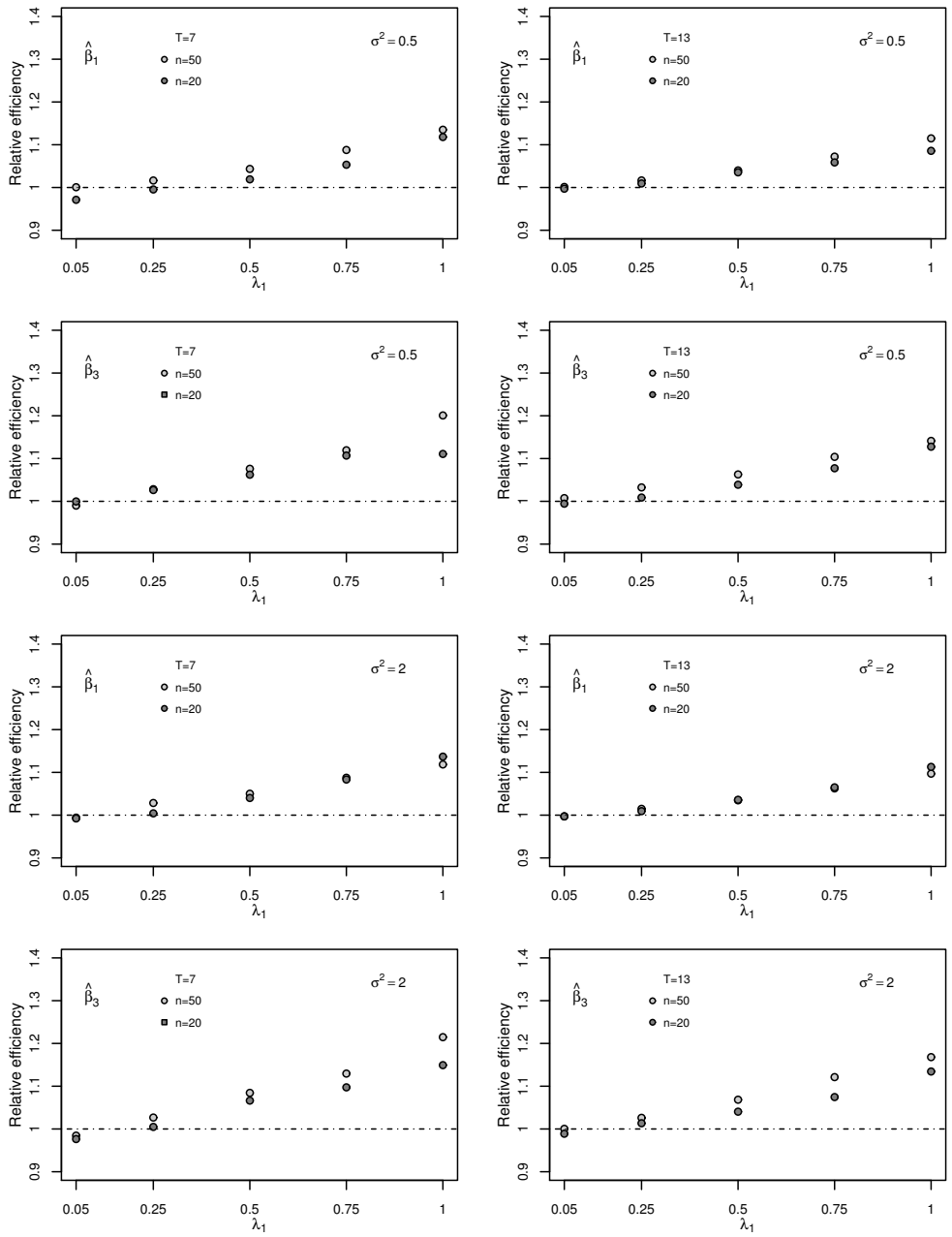


Figure 3: Relative efficiency (see text for definition) of $\hat{\beta}_1$ and $\hat{\beta}_3$ based on 1000 runs using GLM3C and GLMM estimation procedures (T = length of profile on each subject, n = number of subjects, σ^2 = variance of the random effect). Data set on each run has a serial dependence MC1 regulated by λ_1.

3.2. Serial dependence MC2

Under a serial dependence MC2 and to β_1 and β_3 parameters the main conclusions, based on the approaches GLM3C-MC2 and GLMM, are:

(**i**) In both approaches the coverage probabilities are similar when the dependence structure established by (λ_1, λ_2) is low, (0.05,0.05) and (0.25,0.25). The exception occurs when$(\lambda_1, \lambda_2) = (0.05,0.05)$ for $\hat{\beta}_1$ and $\hat{\beta}_3$ when $T = 13$, $n = 20$ and $\sigma^2 = 2$, where the coverage probabilities of GLMM approach are closer to nominal. When (λ_1, λ_2) is greater than (0.25,0.25) and for all the design configurations the coverage probabilities are closer to nominal for the GLM3C approach than for the GLMM approach. This is so much better applied as the length of the profile of each subject increases as well as the dependence structure established by (λ_1, λ_2) (Figures 4–5 and Tables 5–8).

(**ii**) The results of simulation show that when (λ_1, λ_2) is greater than (0.25,0.25) the GLM3C estimators are more efficient than the GLMM estimators and this is so much better applied as (λ_1, λ_2) increases. For values of (λ_1, λ_2) equal to (0.25,0.25) the efficiency of both approaches is similar. When (λ_1, λ_2) is equal to (0.05,0.05) the GLMM estimators are more efficient than the GLM3C in some design configurations (Figure 6 and Tables 5–8).

(**iii**) The estimate bias for the GLMM approach becomes greater than the associated with the GLM3C as (λ_1, λ_2) increases and for all the designs configurations except for $\hat{\beta}_3$ when $T = 13$, $n = 20$ and $\sigma^2 = 2$ (Tables 5–8).

Table 5: Results of the simulation study under a serial dependence MC2 for $n = 20$, $T = 7$.

		GLM3C-MC2					**GLMM**				
		(λ_1, λ_2)					(λ_1, λ_2)				
		$(.05, .05)$	$(.25, .25)$	$(.5, .5)$	$(.75, .75)$	$(1, 1)$	$(.05, .05)$	$(.25, .25)$	$(.5, .5)$	$(.75, .75)$	$(1, 1)$
		$\sigma^2 = 0.5$									
Mean	$\hat{\beta}_1$	0.522	0.514	0.531	0.537	0.546	0.518	0.518	0.550	0.576	0.605
	$\hat{\beta}_3$	1.066	1.130	1.114	1.114	1.123	1.050	1.139	1.152	1.197	1.259
Bias	$\hat{\beta}_1$	0.021	0.014	0.031	0.037	0.046	0.018	0.018	0.050	0.076	0.105
	$\hat{\beta}_3$	0.066	0.130	0.114	0.114	0.123	0.050	0.139	0.152	0.197	0.259
MSE	$\hat{\beta}_1$	0.091	0.106	0.115	0.127	0.140	0.085	0.104	0.123	0.147	0.177
	$\hat{\beta}_3$	0.253	0.318	0.338	0.350	0.372	0.233	0.316	0.358	0.420	0.504
CI	β_1	0.944	0.944	0.946	0.937	0.944	0.946	0.947	0.928	0.905	0.893
	β_3	0.963	0.941	0.946	0.954	0.961	0.974	0.928	0.925	0.914	0.913
		$\sigma^2 = 1$									
Mean	$\hat{\beta}_1$	0.521	0.526	0.514	0.529	0.548	0.515	0.530	0.539	0.571	0.616
	$\hat{\beta}_3$	1.066	1.067	1.130	1.105	1.112	1.060	1.080	1.177	1.198	1.254
Bias	$\hat{\beta}_1$	0.021	0.026	0.014	0.029	0.048	0.015	0.030	0.039	0.071	0.116
	$\hat{\beta}_3$	0.066	0.067	0.130	0.105	0.112	0.060	0.080	0.177	0.198	0.254
MSE	$\hat{\beta}_1$	0.098	0.108	0.114	0.128	0.144	0.092	0.104	0.123	0.147	0.183
	$\hat{\beta}_3$	0.267	0.276	0.364	0.413	0.372	0.263	0.278	0.391	0.482	0.487
CI	β_1	0.944	0.943	0.936	0.938	0.939	0.942	0.944	0.919	0.910	0.892
	β_3	0.955	0.955	0.949	0.936	0.946	0.958	0.946	0.920	0.897	0.904
		$\sigma^2 = 2$									
Mean	$\hat{\beta}_1$	0.507	0.525	0.524	0.496	0.522	0.505	0.536	0.558	0.550	0.601
	$\hat{\beta}_3$	1.054	1.006	0.991	1.112	1.103	1.051	1.033	1.062	1.233	1.272
Bias	$\hat{\beta}_1$	0.007	0.025	0.024	−0.004	0.022	0.005	0.036	0.058	0.050	0.101
	$\hat{\beta}_3$	0.054	0.006	−0.009	0.112	0.103	0.051	0.033	0.062	0.233	0.272
MSE	$\hat{\beta}_1$	0.099	0.098	0.096	0.124	0.132	0.094	0.100	0.109	0.149	0.173
	$\hat{\beta}_3$	0.269	0.266	0.282	0.371	0.385	0.260	0.276	0.307	0.478	0.542
CI	β_1	0.941	0.951	0.956	0.941	0.947	0.944	0.939	0.931	0.902	0.902
	β_3	0.954	0.956	0.952	0.945	0.963	0.957	0.945	0.924	0.904	0.906

Table 6: Results of the simulation study under a serial dependence MC2 for $n = 20$, $T = 13$.

		GLM3C-MC2					**GLMM**				
		(λ_1, λ_2)					(λ_1, λ_2)				
		$(.05, .05)$	$(.25, .25)$	$(.5, .5)$	$(.75, .75)$	$(1, 1)$	$(.05, .05)$	$(.25, .25)$	$(.5, .5)$	$(.75, .75)$	$(1, 1)$
		$\sigma^2 = 0.5$									
Mean	$\hat{\beta}_1$	0.501	0.497	0.501	0.507	0.538	0.501	0.503	0.521	0.540	0.594
	$\hat{\beta}_3$	1.042	1.051	1.064	1.070	1.041	1.044	1.068	1.103	1.141	1.141
Bias	$\hat{\beta}_1$	0.001	−0.003	0.001	0.007	0.038	0.001	0.003	0.021	0.040	0.094
	$\hat{\beta}_3$	0.042	0.051	0.064	0.070	0.041	0.044	0.068	0.103	0.141	0.141
MSE	$\hat{\beta}_1$	0.051	0.057	0.064	0.073	0.081	0.051	0.058	0.071	0.086	0.112
	$\hat{\beta}_3$	0.136	0.155	0.177	0.198	0.210	0.134	0.160	0.199	0.242	0.275
CI	β_1	0.948	0.951	0.947	0.950	0.953	0.946	0.942	0.915	0.876	0.842
	β_3	0.956	0.948	0.956	0.950	0.961	0.955	0.941	0.922	0.895	0.865
		$\sigma^2 = 1$									
Mean	$\hat{\beta}_1$	0.506	0.505	0.535	0.532	0.516	0.507	0.512	0.556	0.569	0.567
	$\hat{\beta}_3$	1.025	1.044	0.944	0.996	1.087	1.030	1.060	0.991	1.07	1.199
Bias	$\hat{\beta}_1$	0.006	0.005	0.035	0.032	0.016	0.007	0.012	0.056	0.069	0.067
	$\hat{\beta}_3$	0.025	0.044	−0.056	−0.004	0.087	0.030	0.060	−0.009	0.077	0.199
MSE	$\hat{\beta}_1$	0.054	0.058	0.071	0.076	0.087	0.053	0.060	0.077	0.091	0.114
	$\hat{\beta}_3$	0.145	0.157	0.156	0.196	0.242	0.144	0.164	0.169	0.231	0.324
CI	β_1	0.942	0.953	0.930	0.939	0.951	0.946	0.938	0.891	0.868	0.847
	β_3	0.936	0.959	0.950	0.948	0.948	0.937	0.948	0.919	0.891	0.854
		$\sigma^2 = 2$									
Mean	$\hat{\beta}_1$	0.536	0.510	0.549	0.549	0.553	0.537	0.516	0.571	0.585	0.608
	$\hat{\beta}_3$	0.981	1.048	0.894	0.882	0.883	0.980	1.063	0.947	0.980	1.024
Bias	$\hat{\beta}_1$	0.036	0.010	0.049	0.049	0.053	0.037	0.016	0.071	0.085	0.108
	$\hat{\beta}_3$	−0.019	0.048	−0.106	−0.118	−0.117	−0.020	0.063	−0.053	−0.020	0.024
MSE	$\hat{\beta}_1$	0.068	0.066	0.072	0.078	0.089	0.067	0.068	0.080	0.094	0.119
	$\hat{\beta}_3$	0.164	0.194	0.168	0.183	0.216	0.159	0.199	0.177	0.205	0.262
CI	β_1	0.939	0.939	0.937	0.948	0.946	0.948	0.928	0.900	0.870	0.835
	β_3	0.933	0.949	0.939	0.948	0.943	0.946	0.946	0.923	0.912	0.880

Table 7: Results of the simulation study under a serial dependence MC2 for $n = 50$, $T = 7$.

		GLM3C-MC2					**GLMM**				
		(λ_1, λ_2)					(λ_1, λ_2)				
		$(.05, .05)$	$(.25, .25)$	$(.5, .5)$	$(.75, .75)$	$(1, 1)$	$(.05, .05)$	$(.25, .25)$	$(.5, .5)$	$(.75, .75)$	$(1, 1)$
		$\sigma^2 = 0.5$									
Mean	$\hat{\beta}_1$	0.506	0.495	0.496	0.513	0.532	0.506	0.509	0.528	0.567	0.612
	$\hat{\beta}_3$	1.018	1.070	1.077	1.052	1.045	1.017	1.098	1.142	1.163	1.210
Bias	$\hat{\beta}_1$	0.006	−0.005	−0.003	0.013	0.032	0.006	0.009	0.028	0.067	0.112
	$\hat{\beta}_3$	0.018	0.070	0.077	0.052	0.045	0.017	0.098	0.142	0.163	0.210
MSE	$\hat{\beta}_1$	0.037	0.035	0.039	0.043	0.049	0.037	0.036	0.043	0.055	0.074
	$\hat{\beta}_3$	0.101	0.111	0.123	0.115	0.124	0.099	0.117	0.146	0.157	0.196
CI	β_1	0.953	0.956	0.951	0.946	0.950	0.947	0.946	0.933	0.907	0.874
	β_3	0.947	0.952	0.941	0.953	0.957	0.943	0.924	0.912	0.909	0.881
		$\sigma^2 = 1$									
Mean	$\hat{\beta}_1$	0.503	0.512	0.520	0.517	0.520	0.505	0.527	0.554	0.572	0.598
	$\hat{\beta}_3$	1.049	1.045	1.044	1.063	1.070	1.050	1.072	1.114	1.176	1.237
Bias	$\hat{\beta}_1$	0.003	0.012	0.020	0.017	0.020	0.005	0.027	0.054	0.072	0.098
	$\hat{\beta}_3$	0.049	0.045	0.044	0.063	0.070	0.050	0.072	0.114	0.176	0.237
MSE	$\hat{\beta}_1$	0.036	0.040	0.044	0.043	0.049	0.036	0.042	0.051	0.057	0.070
	$\hat{\beta}_3$	0.101	0.099	0.123	0.133	0.151	0.098	0.103	0.141	0.174	0.229
CI	β_1	0.952	0.943	0.947	0.961	0.949	0.950	0.926	0.922	0.910	0.885
	β_3	0.955	0.948	0.938	0.944	0.937	0.945	0.932	0.914	0.897	0.866
		$\sigma^2 = 2$									
Mean	$\hat{\beta}_1$	0.509	0.506	0.498	0.510	0.524	0.511	0.522	0.535	0.568	0.609
	$\hat{\beta}_3$	1.062	1.047	1.065	1.049	1.038	1.063	1.076	1.144	1.171	1.215
Bias	$\hat{\beta}_1$	0.009	0.006	−0.002	0.010	0.024	0.011	0.022	0.035	0.068	0.109
	$\hat{\beta}_3$	0.062	0.047	0.065	0.049	0.038	0.063	0.076	0.144	0.171	0.215
MSE	$\hat{\beta}_1$	0.040	0.039	0.042	0.046	0.049	0.040	0.041	0.048	0.059	0.075
	$\hat{\beta}_3$	0.105	0.103	0.117	0.129	0.140	0.102	0.107	0.145	0.170	0.215
CI	β_1	0.949	0.941	0.938	0.955	0.966	0.943	0.929	0.917	0.909	0.883
	β_3	0.951	0.953	0.951	0.953	0.951	0.947	0.939	0.909	0.902	0.873

Table 8: Results of the simulation study under a serial dependence MC2 for $n = 50$, $T = 13$.

		GLM3C-MC2					**GLMM**				
		(λ_1, λ_2)					(λ_1, λ_2)				
		$(.05, .05)$	$(.25, .25)$	$(.5, .5)$	$(.75, .75)$	$(1, 1)$	$(.05, .05)$	$(.25, .25)$	$(.5, .5)$	$(.75, .75)$	$(1, 1)$
		$\sigma^2 = 0.5$									
Mean	$\hat{\beta}_1$	0.497	0.499	0.503	0.506	0.510	0.498	0.507	0.523	0.541	0.564
	$\hat{\beta}_3$	1.042	1.044	1.041	1.038	1.044	1.045	1.062	1.083	1.113	1.160
Bias	$\hat{\beta}_1$	−0.003	−0.001	0.003	0.006	0.010	−0.001	0.007	0.023	0.041	0.064
	$\hat{\beta}_3$	0.042	0.044	0.041	0.038	0.044	0.045	0.062	0.083	0.113	0.160
MSE	$\hat{\beta}_1$	0.022	0.024	0.028	0.030	0.035	0.022	0.025	0.031	0.038	0.049
	$\hat{\beta}_3$	0.058	0.065	0.073	0.079	0.089	0.059	0.069	0.085	0.104	0.138
CI	β_1	0.950	0.953	0.952	0.949	0.949	0.953	0.934	0.901	0.884	0.836
	β_3	0.944	0.945	0.943	0.938	0.948	0.941	0.919	0.900	0.867	0.818
		$\sigma^2 = 1$									
Mean	$\hat{\beta}_1$	0.498	0.506	0.506	0.499	0.507	0.500	0.516	0.527	0.536	0.562
	$\hat{\beta}_3$	1.035	1.025	1.027	1.056	1.044	1.038	1.043	1.069	1.129	1.160
Bias	$\hat{\beta}_1$	−0.002	0.006	0.006	−0.001	0.007	0.000	0.016	0.027	0.036	0.062
	$\hat{\beta}_3$	0.035	0.025	0.027	0.056	0.044	0.038	0.043	0.069	0.129	0.160
MSE	$\hat{\beta}_1$	0.022	0.025	0.028	0.028	0.033	0.022	0.026	0.031	0.034	0.045
	$\hat{\beta}_3$	0.051	0.060	0.068	0.071	0.083	0.052	0.063	0.078	0.094	0.128
CI	β_1	0.950	0.941	0.943	0.952	0.954	0.940	0.922	0.901	0.883	0.856
	β_3	0.954	0.953	0.944	0.956	0.942	0.953	0.926	0.889	0.874	0.828
		$\sigma^2 = 2$									
Mean	$\hat{\beta}_1$	0.492	0.500	0.503	0.503	0.513	0.493	0.509	0.524	0.539	0.568
	$\hat{\beta}_3$	1.048	1.036	1.038	1.041	1.037	1.051	1.054	1.079	1.114	1.148
Bias	$\hat{\beta}_1$	−0.008	0.000	0.003	0.003	0.013	−0.007	0.009	0.024	0.039	0.068
	$\hat{\beta}_3$	0.048	0.036	0.038	0.041	0.037	0.051	0.054	0.079	0.114	0.148
MSE	$\hat{\beta}_1$	0.024	.030	0.033	0.038	0.042	0.024	0.031	0.037	0.045	0.055
	$\hat{\beta}_3$	0.060	0.071	0.079	0.088	0.097	0.061	0.076	0.091	0.114	0.143
CI	β_1	0.952	0.936	0.939	0.943	0.948	0.951	0.928	0.900	0.876	0.848
	β_3	0.944	0.951	0.956	0.959	0.950	0.942	0.935	0.915	0.880	0.844

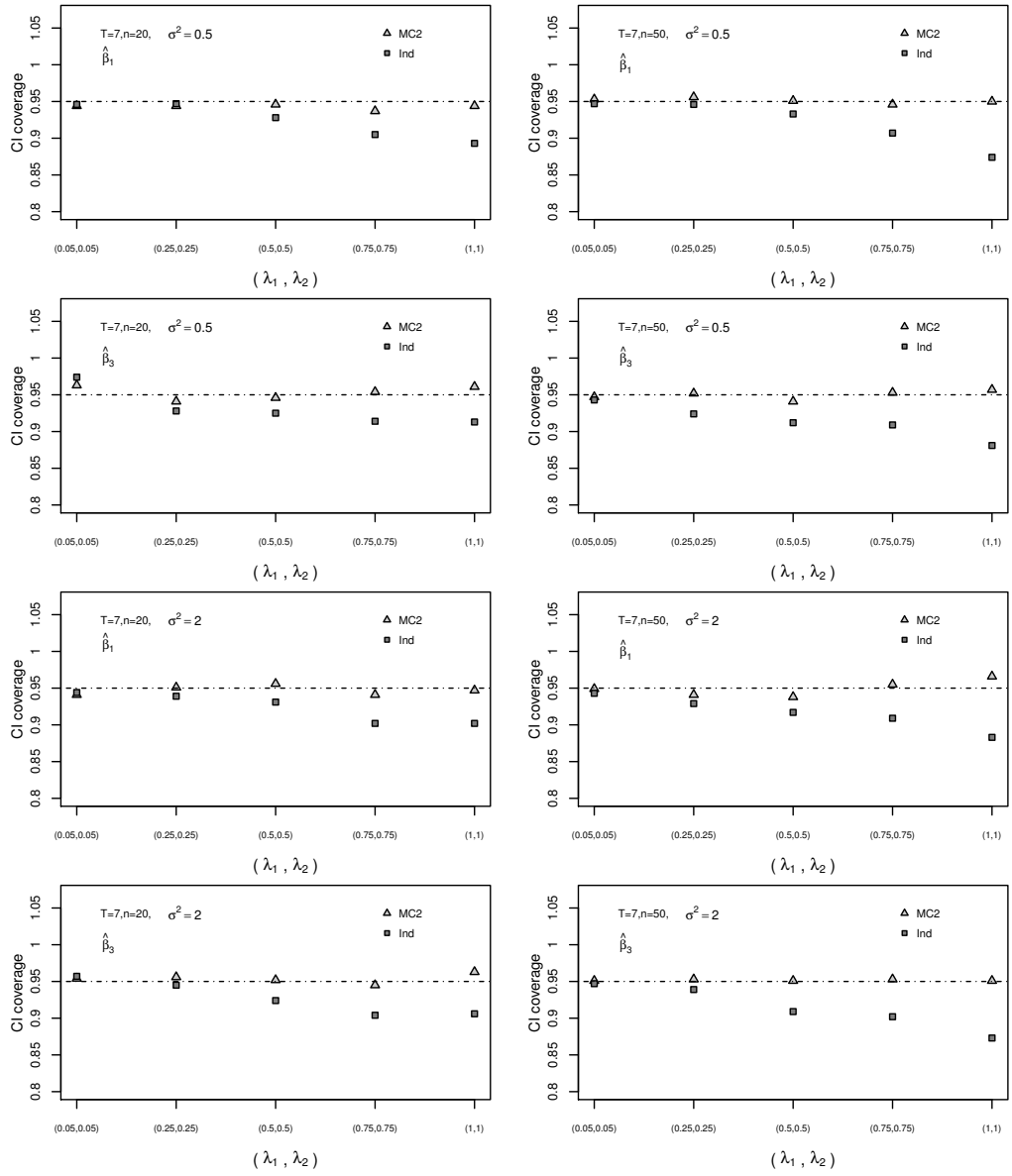


Figure 4: Coverage probabilities of nominal 95% confidence intervals (CI coverage) for β_1 ($\hat{\beta}_1$) and β_3 ($\hat{\beta}_3$) based on 1000 runs using GLM3C and GLMM estimation procedures (T = length of profile on each subject, n = number of subjects, σ^2 = variance of the random effect). Data set on each run has a serial dependence MC2 regulated by $\lambda = (\lambda_1, \lambda_2)$. Coding for estimation procedures: MC2 (GLM3C) and Ind (GLMM).

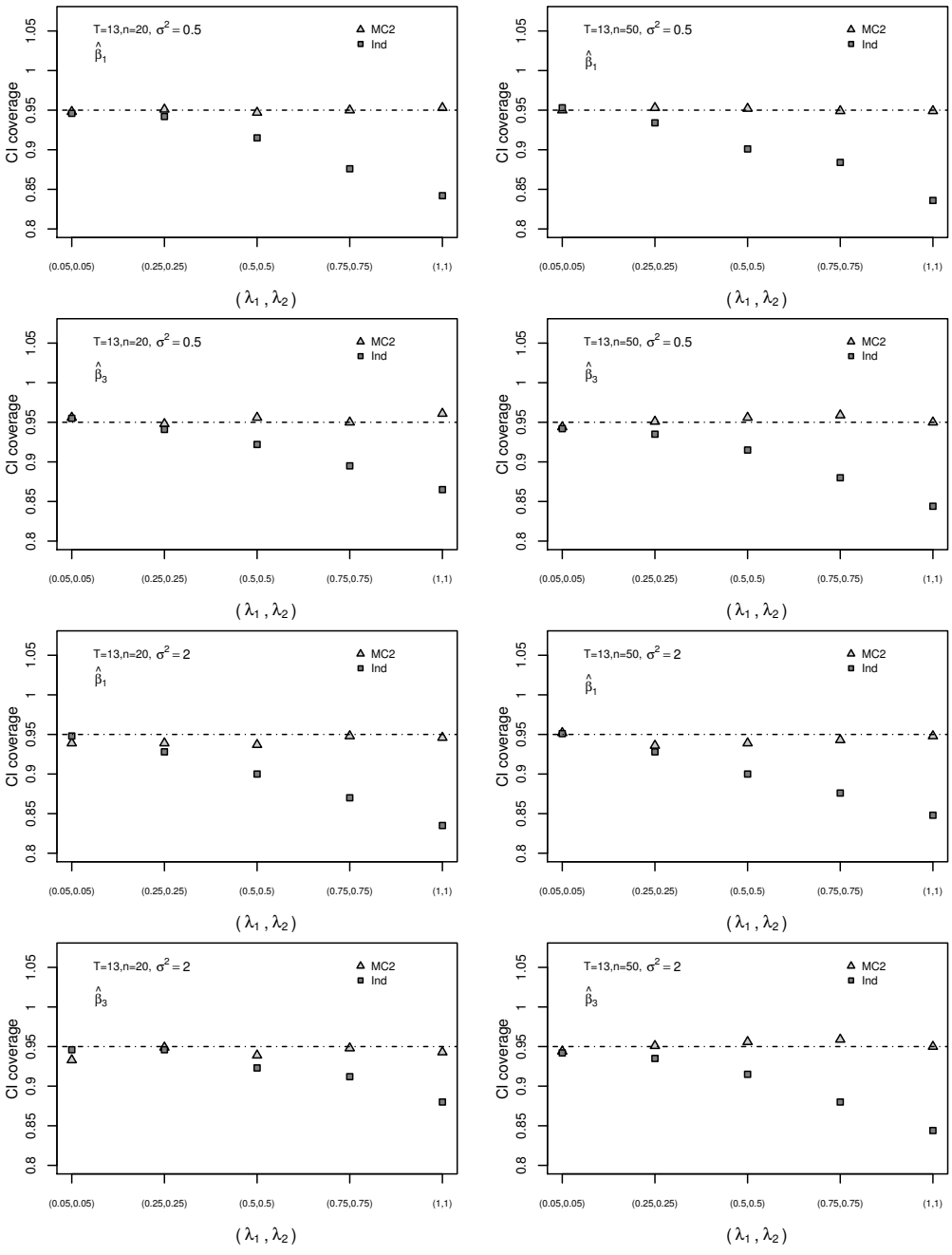


Figure 5: Coverage probabilities of nominal 95% confidence intervals (CI coverage) for β_1 ($\hat{\beta}_1$) and β_3 ($\hat{\beta}_3$) based on 1000 runs using GLM3C and GLMM estimation procedures (T = length of profile on each subject, n = number of subjects, σ^2 = variance of the random effect). Data set on each run has a serial dependence MC2 regulated by $\lambda = (\lambda_1, \lambda_2)$. Coding for estimation procedures: MC2 (GLM3C) and Ind (GLMM).

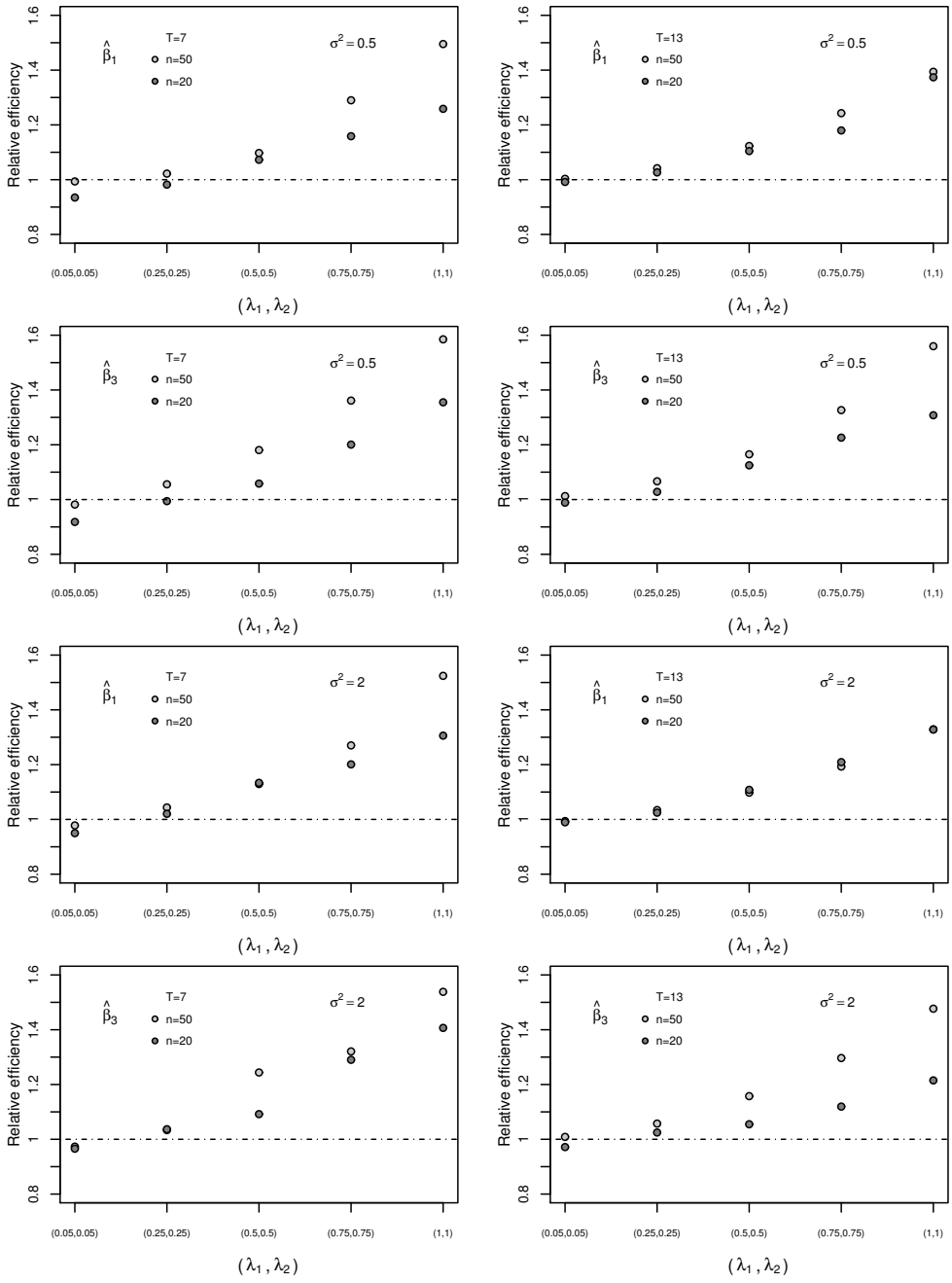


Figure 6: Relative efficiency (see text for definition) of $\hat{\beta}_1$ and $\hat{\beta}_3$ based on 1000 runs using GLM3C and GLMM estimation procedures (T = length of profile on each subject, n = number of subjects, σ^2 = variance of the random effect). Data set on each run has a serial dependence MC2 regulated by $\lambda = (\lambda_1, \lambda_2)$.

4. ANALYSIS OF CONTRACEPTING WOMEN DATA

The key results of the simulation study are illustrated, using data from a longitudinal clinical trial of contracepting women given in [7]. In this trial, and following their description, women received an injection of either 100 mg or 150 mg of depot-medroxyprogesterone acetate (DMPTA) on the day of randomization and three additional injections at 90-day intervals. There was a final follow-up visit 90 days after the fourth injection. The outcome of interest is a binary response indicating whether the i^{th} woman experienced amenorrhea ($Y_{ij} = 1$) in the j^{th} four successive three-month intervals, or not ($Y_{ij} = 0$). A feature of this clinical trial is that there was substantial dropout (17% dropped out after receiving one injection of DMPA, 13% dropped out after receiving only two injections, and 7% dropped out after receiving three injections).

The mixed effects logistic model proposed by [7]

$$\begin{aligned}\text{logit } [E(Y_{ij}|b_i)] \;=\; & \beta_0 + \beta_1 time_{ij} + \beta_2 time_{ij}^2 + \beta_3 (time_{ij} \times dose_i) \\ & + \beta_4 (time_{ij}^2 \times dose_i) + b_i\end{aligned}$$

with $j = 1, ..., 4$, $time = 1, 2, 3, 4$ and *dose* a binary variable taking the value 1 if the i-th woman is randomized to 150mg of DMPA and 0 otherwise, was fitted to data with different dependence structures:

(**i**) Independence (Model I).

(**ii**) Serial dependence MC1 (Model II).

(**iii**) Serial dependence MC2 (Model III).

Models I correspond to the GLMM approach (model fitted by [7]), Models II and III correspond to the GLM3C approach. The analysis of all models was performed using the `bild` function of the `R` package `bild` with the dependence argument sated to `indR` to Model I, `MC1R` and `MC2R`, respectively to Models II and III.

Tables 9 and 10 display the results of fitting the different models to data. Table 9 reports the log-likelihood, the change in deviance with corresponding p-values. The estimated values of the parameters, as well as their standard errors, t-ratio and corresponding p-values are given in Table 10.

The first step of the analysis is to choose the appropriate serial dependence to account correlation between successive observation of the same subject. The results of Table 10 show, among other things, that the estimates of λ_1 and λ_2 in MC2 model (Model III) as well as the estimate of λ_1 in MC1 model (Model II) point strongly to a first order serial dependence. The change of deviance between this two models, compared with the χ_1^2 reference distribution, produces a p-value = 0.3467 (Table 9) confirming that there is no significant difference between this

two models at 5% level. To explore further this point Model II was compare to Model I, which assume independence between successive observation of the same subject. The change of deviance between this two models, compared with the χ^2_1 reference distribution, produces a p-value= 0.0001 (Table 9) and so Model II with a serial dependence MC1 is significantly preferable to Model I.

Table 9: Log-likelihood and change in deviance between models.

Model	**LogL**	**Δ D**	**p-value**
I	−1937.54		
II	−1930.108	14.866	0.0001
II	−1930.108		
III	−1929.665	0.885	0.3467

The results displayed in Table 10 also show that the model with the appropriate serial dependence (MC1 model–Model II) produce smaller standard errors, as point out in the simulation study, as well as a decrease in the value of the estimative of σ^2 face a more complex serial dependence model as remarked by Pinheiro and Bates (2000) ([17]).

Table 10: Parameters estimates, Standard errors, t-ratio and p-value for models I, II and III.

Model	**Parameter**	**Estimate**	**SE**	**t-ratio**	**p-value**
I	β_0	−3.799	0.305	−12.471	0.0000
	β_1	1.131	0.268	4.221	0.0000
	β_2	−0.042	0.055	−0.763	0.4457
	β_3	0.562	0.192	2.932	0.0034
	β_4	−0.109	0.050	−2.206	0.0274
	σ^2	5.030			
II	β_0	−3.443	0.304	−11.328	0.0000
	β_1	1.033	0.247	4.188	0.0000
	β_2	−0.039	0.050	−0.781	0.4346
	β_3	0.522	0.177	2.943	0.0033
	β_4	0.105	0.177	2.943	0.0234
	λ_1	0.744	0.226	3.293	0.0009
	σ^2	3.598			
III	β_0	−3.384	0.397	−8.524	0.0000
	β_1	1.014	0.253	4.004	0.0000
	β_2	−0.038	0.049	−0.781	0.4349
	β_3	0.516	0.178	2.904	0.0037
	β_4	−0.105	0.046	−2.262	0.0237
	λ_1	0.820	0.397	2.068	0.0387
	λ_2	0.092	0.398	0.230	0.8178
	σ^2	3.376			

The model fitted by [7] corresponds to Model I. The estimated values of the parameters, as well as, their standard errors reported in Table 10 are in close agreement to those obtained by [7]. To fit Model I [7] used the `PROC NLMIXED` procedure in `SAS` and the estimation was based on 50-point adaptive Gaussian quadrature.

5. FINAL REMARKS

This paper is concerned with the asses of performance of the GLM3C and GLMM approaches both implemented in `R` package `bild` for the analysis of longitudinal binary data. The GLM3C approach seems to be preferable to GLMM in the situations considered by checking that its performance is so much better the higher the serial correlation between observations of the same subject, regardless of the number of subjects involved in the study, the length of their profile or the variance of the random effect. In spite of the use of the adaptive Gaussian quadrature method the users may be aware that this method needs careful handling to ensure converge even in simple random-effects models for categorical outcome data as referred in [15].

The results pointed out in the simulation study are illustrated in the example analyzed where a MC1 model was need to account dependence between successive observations of the same subject. The program codes for analysing the data set are available under request from the authors.

Finally, the `R` package `bild` allows the practitioners to choose the serial dependence adequate to use for their data at hand.

ACKNOWLEDGMENTS

This work has been partially funded by FCT – Fundação Nacional para a Ciência e a Tecnologia, Portugal, through the project UID/MAT/00006/2013. The authors thank the Referee, the Associate Editor and the Editor-in-Chief for their comments which have led to significant improvement of this paper.

REFERENCES

[1] Azzalini, A. (1994). Logistic Regression for Autocorrelated Data with Application to Repeated Measures, *Biometrika*, **81**(4), 767–775.

[2] Bates, D.; Maechler, M. and Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes *R package* version 0.999375-42.
URL: `http://CRAN.R-project.org/package=lme4`

[3] Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 9–25.

[4] Fitzmaurice, G.M. and Laird, N.M. (1993). A likelihood-based method for analysing longitudinal binary responses, *Biometrika*, **80**, 141–151.

[5] Fitzmaurice, G.M.; Laird, N.M. and Rotnitzky, A.G. (1993). Regression models for discrete longitudinal responses, *Statistical Science*, **8**(3), 141–151.

[6] Fitzmaurice, G.M.; Laird, N.M. and Lipsitz, S.R. (1994). Analyzing incomplete longitudinal binary responses: a likelihood-based approach, *Biometrics*, **50**, 601–612.

[7] Fitzmaurice, G.M.; Laird, N.M. and Ware, J.H. (2004). *Applied Longitudinal Analysis*, John Wiley & Sons, New York.

[8] Gonçalves, M.H. (2002). *Likelihood methods for discrete longitudinal data*, PhD thesis, University of Lisbon.

[9] Gonçalves, M.H. and Azzalini, A. (2008). Using Markov chains for marginal modelling of binary longitudinal data in an exact likelihood approach, *Metron*, **LXVI**(1), 157–181.

[10] Gonçalves, M.H.; Cabral, M.S. and Azzalini, A. (2012). The R package bild for the analysis of binary longitudinal data, *Journal of statistical software*, **46**(9), 1–17.

[11] Gonçalves, M.H.; Cabral, M.S. and Azzalini, A. (2013). bild: A Package for Binary Longitudinal Data, Version 1.1-5.
URL: `http://CRAN.R-project.org/package=bild`

[12] Jansen, I.; Beunckens, C.; Molenberghs, G.; Verbeke, G. and Mallinckrodt, C. (2006). Analyzing incomplete discrete longitudinal clinical trials data, *Statistical Science*, **21**(1), 52–69.

[13] Laird, N. (1988). Missing data in longitudinal studies, *Statistics in Medicine*, **7**, 505–315.

[14] Lipsitz, S.R.; Molenberghs, G.; Fitzmaurice, G.M. and Ibrahim, J. (2000). GEE with Gaussian estimation of the correlations when data are incomplete, *Biometrics*, **56**, 528–536.

[15] Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example, *Applied Statistics*, **50**, 325–335.

[16] Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*, John Wiley & Sons, New York.

[17] Pinheiro, J.C. and Bates, D.M. (2004). *Mixed effects models in S and S-Plus*, Springer-Verlag, New York.

[18] R DevelopmentCore Team (2013). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: `http://www.R-project.org`

REVSTAT – Statistical Journal
Volume 16, Number 4, October 2018, 515–536

SPATIO-TEMPORAL ANALYSIS OF REGIONAL UNEMPLOYMENT RATES: A COMPARISON OF MODEL BASED APPROACHES

Authors: Soraia Pereira
– Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
sapereira@fc.ul.pt

Feridun Turkman
– Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
kfturkman@fc.ul.pt

Luís Correia
– Instituto Nacional de Estatística, Lisboa, Portugal
luis.correia@ine.pt

Received: April 2016 Revised: November 2016 Accepted: November 2016

Abstract:

- This study aims to analyze the methodologies that can be used to estimate the total number of unemployed, as well as the unemployment rates for 28 regions of Portugal, designated as NUTS III regions, using model based approaches as compared to the direct estimation methods currently employed by INE (National Statistical Institute of Portugal). Model based methods, often known as small area estimation methods (Rao, 2003), "borrow strength" from neighbouring regions and in doing so, aim to compensate for the small sample sizes often observed in these areas. Consequently, it is generally accepted that model based methods tend to produce estimates which have lesser variation. Other benefit in employing model based methods is the possibility of including auxiliary information in the form of variables of interest and latent random structures. This study focuses on the application of Bayesian hierarchical models to the Portuguese Labor Force Survey data from the 1st quarter of 2011 to the 4th quarter of 2013. Three different data modeling strategies are considered and compared: Modeling of the total unemployed through Poisson, Binomial and Negative Binomial models; modeling of rates using a Beta model; and modeling of the three states of the labor market (employed, unemployed and inactive) by a Multinomial model. The implementation of these models is based on the *Integrated Nested Laplace Approximation* (INLA) approach, except for the Multinomial model which is implemented based on the method of Monte Carlo Markov Chain (MCMC). Finally, a comparison of the performance of these models, as well as the comparison of the results with those obtained by direct estimation methods at NUTS III level are given.

Key-Words:

- *unemployment estimation; model based methods; Bayesian hierarchical models.*

1. INTRODUCTION

The calculation of official estimates of the labor market that are published quarterly by the INE is based on a direct method from the sample of the Portuguese Labor Force Survey. These estimates are available at national level and NUTS II regions of Portugal. NUTS is the classification of territorial units for statistics (see Appendix for a better understanding). Currently, as established by Eurostat, knowledge of the labor market requires reliable estimates for the total of unemployed people and the unemployment rate at more disaggregated levels, particularly at NUTS III level. However, due to the small size of these areas, there is insufficient information on some of the variables of interest to obtain estimates with acceptable accuracy using the direct method.

In this sense, and because increasing the sample size imposes excessive costs, we intend to study alternative methods with the aim of getting more accurate estimates for these regions. In fact, the accuracy of the estimates obtained in this context is deemed very important since it directly affects the local policy actions.

This issue is part of the *small area estimation*. Rao (2003) provides a good theoretical introduction to this problem and discusses some estimation techniques based on mixed generalized models. Pfeffermann (2002), Jiang & Lahiri (2006a, 2006b) make a good review of developments to date.

This has been an area in full development and application, especially after the incorporation of spatial and temporal random effects, which brought a major improvement in the estimates produced. Choundry & Rao (1989), Rao & Yu (1994), Singh *et al.* (2005) and Lopez-Vizcaino *et al.* (2015) are responsible for some of these developments. Chambers *et al.* (2016) give alternative semiparametric methods based on M-quantile regression.

Datta & Ghosh (1991) use a Bayesian approach for the estimation in small areas. One advantage of using this approach is the flexibility in modeling different types of variables of interest and different structures in the random effects using the same computational methods.

Recently, there has been considerable developments on space-time Bayesian hierarchical models employed in small area estimation within the context of disease (Best *et al.*, 2005). In this paper, we explore the application of these models and adopt them for the estimation of unemployment in the NUTS III regions, using data from the Portuguese Labor Force Survey from the 1st quarter of 2011 to the 4th quarter of 2013.

We consider three different modeling strategies: the modeling of the total number of unemployed people through the Poisson, Binomial, and Negative Binomial models; modeling the unemployment rate using a Beta model; and the simultaneous modeling of the total of the three categories of the labor market (employment, unemployment and inactivity) using a Multinomial model.

2. DATA

The region under study (Portugal Continental) is partitioned into 28 NUTS III regions, indexed by $j = 1, ..., 28$. We did not include the autonomous regions because they coincide with the NUTS II regions for which estimates are already available with acceptable accuracy.

We use the Portuguese Labor Force Survey data from the 1st quarter of 2011 to the 4th quarter of 2013 in order to produce accurate estimates for the labor market indicators in the last quarter. Each quarter is denoted by $t = 1, ..., 12$. We did not use more recent data because there was a change in the sampling design during 2014 and that could affect the temporal analysis.

We are interested in the total unemployed population, and the unemployment rate of the population by NUTS III regions, which is denoted by Y_{jt} and R_{jt}. We denote the respective sample values by y_{jt} and r_{jt}. The unemployment rate is given by the ratio of active people who are unemployed, as defined by the European regulation of the Labor Force Survey.

The models developed to make estimation in small areas gain special importance with the inclusion of variables of interest, which we call covariates. In this study, the covariates are divided into 5 groups: population structure, economy, labor market, companies and type of economic activity. Some of these covariates are regional and are static in time whereas others are available per quarter and thus are also of dynamic nature. We will make the distinction and classify these sets of covariates into regional, temporal and spatio-temporal covariates. These selected covariates are as follows:

a) Population structure:
- **a.1**) Proportion of individuals in the sample of the Labor Force Survey that are female and aged between 24 and 34 years (SA6, regional and quarterly);
- **a.2**) Proportion of individuals in the sample of the Labor Force Survey that are female and over 49 years (SA8, regional and quarterly);

b) Economy:
- **b.1**) Gross domestic product per capita (GDP, quarterly);

c) Labor market:
- **c.1**) Proportion of unemployed people registered in the employment centers (IEFP, regional and quarterly);

d) Companies:
- **d.1**) Number of enterprises per 100 inhabitants (regional);

e) Type of economic activity:

- **e.1**) Proportion of population employed in the primary sector of activity (regional);
- **e.2**) Proportion of population employed in the secondary sector of activity (regional).

Figure 1 shows the evolution of the unemployment rate observed in the sample from the Portuguese Labor Force Survey from the 1st quarter of 2011 to the 4th quarter of 2013 in each of the 28 NUTS III . The bold represents the average unemployment rate. We can see that for all regions there was a slight increase in the unemployment rate during this period.

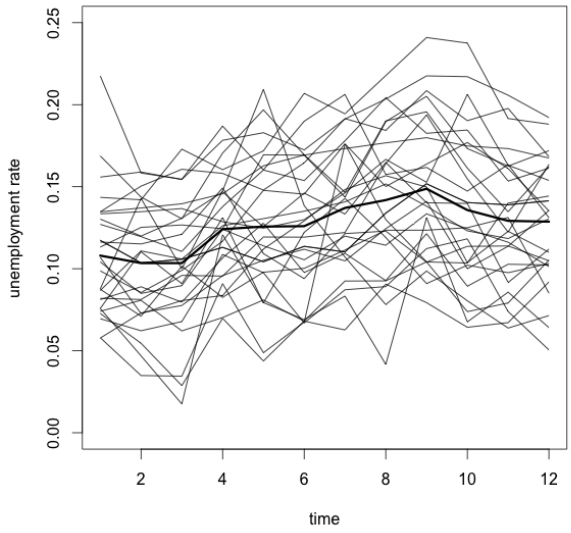


Figure 1: Unemployment rate observed in the sample from the Portuguese Labor Force Survey from the 1st quarter of 2011 to the 4th quarter of 2013 in each of the 28 NUTS III.

The map in Figure 2 shows the spatial and temporal distribution of the unemployment rate observed in the sample of Portuguese Labor Force Survey during the period under study. As we can see, this map suggests the existence of spatial and temporal dependence structures in the observed data.

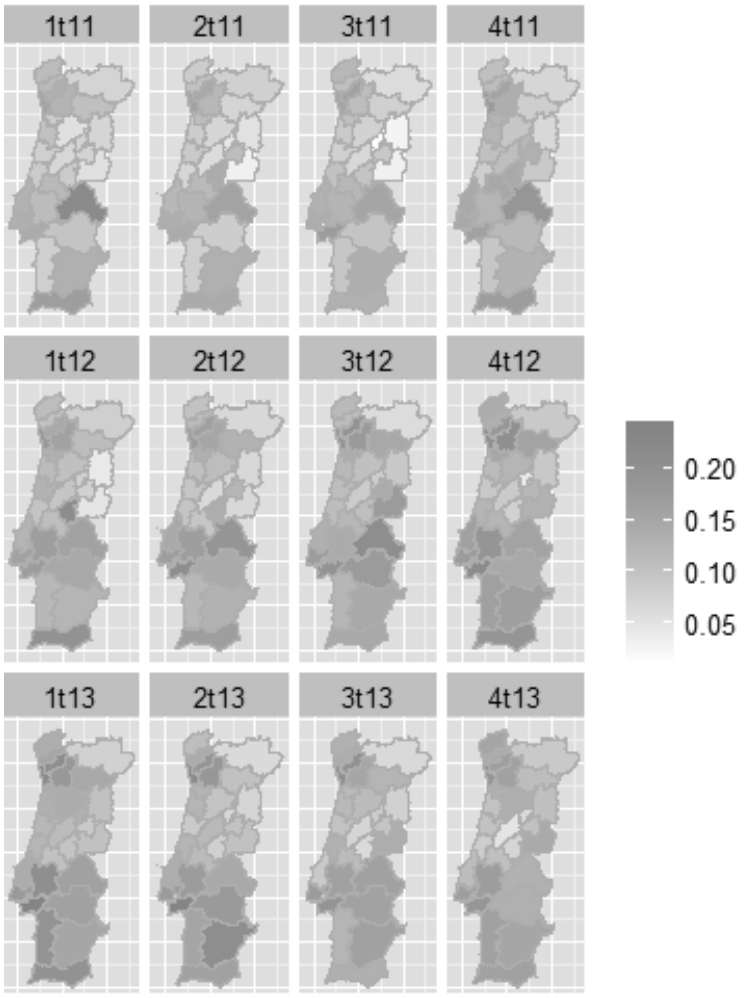


Figure 2: Spatial and temporal distribution of the unemployment rate observed in the sample of Portuguese Labor Force Survey.

3. BAYESIAN MODELS FOR COUNTS AND PROPORTIONS

In this problem we are interested in estimating the effect of selected variables on the number of unemployed individuals and the unemployment rate, taking into account the temporal and spatial correlations.

One of the most general and useful ways of specifying this problem is to employ hierarchical generalized linear model set up, in which the data are linked to covariates and spatial-temporal random effects through an appropriately chosen likelihood and a link function which is linear on the covariates and the random effects.

We denote the vector of designated regional covariates by $\boldsymbol{x_j} = (x_{1j}, x_{2j}, x_{3j})$, the temporal covariates by x_t and the vector of spatio-temporal covariates by $\boldsymbol{x_{jt}} = (x_{1jt}, x_{2jt}, x_{3jt})$.

While modeling unemployment numbers, we generically assume that

$$y_{jt}\,|\,\mu_{jt} \;\sim\; \pi(y_{jt}\,|\,\mu_{jt})\,, \qquad j = 1, ..., 28\,, \quad t = 1, ..., 12\,,$$

where π is a generic probability mass function. We look at this model considering specific probability mass functions, such as Poisson and Binomial, among others. The state parameters μ_{jt} depend on covariates and on structured and unstructured random factors through appropriate link functions.

The unemployment rate is also hierarchically modeled in a similar way. We assume that

$$r_{jt}\,|\,\theta_{jt} \;\sim\; g(r_{jt}\,|\,\theta_{jt})\,, \qquad j = 1, ..., 28\,, \quad t = 1, ..., 12\,,$$

where g is a properly chosen probability density function and θ_{jt} are the state parameters.

In the following sections we look at different variations of these hierarchical structures with different link functions.

Let us consider h, the chosen link function which depends on the assumed model for the data. We assume $\eta_{jt} = h(\mu_{jt})$ for the modeling of the total and $\eta_{jt} = h(\theta_{jt})$ for the modeling of the rates. For each model, we consider the following linear predictor

$$\eta_{jt} \;=\; \mathit{offset}_{jt} + \alpha_0 + \boldsymbol{x_j'}\boldsymbol{\alpha} + x_t'\beta + \boldsymbol{x_{jt}'}\boldsymbol{\gamma} + w_{jt} + \epsilon_{jt}\,, \qquad j = 1, ..., 28\,, \quad t = 1, ..., 12\,, \tag{3.1}$$

where offset_{jt} are constants that can be included in the linear predictor during adjustment. The vectors $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$, β and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)$ correspond respectively to vectors of the covariates coefficients $\boldsymbol{x_j}$, x_t and $\boldsymbol{x_{jt}}$. Components ϵ_{jt} represent unstructured random effects, which assume

$$\epsilon_{jt} \;\sim\; N(0, \sigma_\epsilon^2)\,,$$

and the components w_{jt} represent the structured random effects that can be written as $w_{jt} = w_{1j} + w_{2t}$ where $\boldsymbol{w_1}$ is modeled as a *intrinsic conditional autoregressive* (ICAR) process proposed by Besag *et al.* (1991) and $\boldsymbol{w_2}$ is modeled as a first order *random walk* (AR (1)). Blangiardo *et al.* (2013) succinctly describe both the ICAR and AR (1) processes.

$$\boldsymbol{w_1}\,|\,\tau_{w_1} \sim ICAR(\tau_{w_1})\,,$$
$$\boldsymbol{w_2}\,|\,\tau_{w_2} \sim AR(1)\,.$$

We assume the following prior distributions for the regression parameters

$$\begin{aligned}
&\alpha_0 \sim N(0, 10^6)\,, \\
&\alpha_i \sim N(0, 10^6)\,, \qquad i = 1, 2, 3\,, \\
&\beta \sim N(0, 10^6)\,, \\
&\gamma_i \sim N(0, 10^6)\,, \qquad i = 1, 2, 3\,.
\end{aligned}$$

For the hyperparameters we assume

$$\begin{aligned}
\log \tau_\epsilon &\sim \log Gamma(1, 0.0005)\,, \\
\log \tau_{w_1} &\sim \log Gamma(1, 0.0005), \\
\log \tau_{w_2} &\sim \log Gamma(1, 0.0005)\,.
\end{aligned}$$

We assume the following models for the distribution of the observed data: Poisson, Binomial, and Negative Binomial for the total of unemployed, Beta for the unemployment rate and Multinomial for the total of the three states of the labor market (employment, unemployment and inactivity).

3.1. Poisson model

This is perhaps the most frequently used model for counting data in small areas, especially in epidemiology. If we consider that μ_{jt} is the mean of the total number of unemployed people, we can assume that

$$y_{jt}|\mu_{jt} \sim Poisson(\mu_{jt})\,, \qquad j = 1, ..., 28\,, \quad t = 1, ..., 12\,.$$

Therefore

$$p(y_{jt}|\mu_{jt}) = \mu_{jt}^{y_{jt}} \exp(-\mu_{jt})/y_{jt}!\,, \qquad y_{jt} = 0, 1, 2...$$

In this case, the link function is the logarithmic function ($\log = h$). The NUTS III regions have different sample dimensions, so the variation of the total unemployment is affected. To remove this effect, we need to add an *offset* term, which is given by the number of individuals in the sample in each NUTS III region.

3.2. Negative Binomial model

The Negative Binomial model may be used as an alternative to the Poisson model, especially when the sample variance is much higher than the sample mean. When this happens, we say that there is over-dispersion in the data. In this case, we can assume that

$$y_{jt}|\mu_{jt}, \phi \sim Negative\,Binomial(\mu_{jt}, \phi)\,, \qquad j = 1, ..., 28\,, \quad t = 1, ..., 12\,.$$

The probability mass function is given by

$$p(y_{jt}|\mu_{jt},\phi) \;=\; \frac{\Gamma(y_{jt}+\phi)}{\Gamma(\phi).y_{jt}!}\cdot\frac{\mu_{jt}^{y_{jt}}\cdot\phi^{\phi}}{(\mu_{jt}+\phi)^{y_{jt}+\phi}}\,, \qquad y_{jt}=0,1,2...$$

where $\Gamma(\cdot)$ is the gamma function.

The most convenient way to connect μ_{jt} to the linear predictor is through the $\log\frac{\mu_{jt}}{\mu_{jt}+\phi}$. Also in this case, the term *offset* described in the Poisson model is considered.

3.3. Binomial model

When measuring the total unemployed, we may also consider that there is a finite population in the area j. In this case, we assume that this population is the number of active individuals in the area j, which is denoted by m_{jt}, assuming that it is fixed and known. We can then consider a Binomial model for the total number of unemployed given the observed active population. So, given the population unemployment rate R_{jt},

$$y_{jt}|m_{jt},R_{jt} \sim \mathit{Binomial}(m_{jt},R_{jt})\,, \qquad j=1,...,28\,, \quad t=1,...,12\,,$$

which means that

$$p(y_{jt}|m_{jt},R_{jt}) = \binom{m_{jt}}{y_{jt}} R_{jt}^{y_{jt}}(1-R_{jt})^{m_{jt}-y_{jt}}\,, \qquad y_{jt}=0,1,...,m_{jt}\,, \quad t=1,...,12\,.$$

In this case, the most usual link function is the logit function given by $\log(R_{jt}/(1-R_{jt}))$.

We expect that the fit of this model will be close to the fit of the Poisson model in the regions with a big number of active people and a small unemployment rate.

3.4. Beta model

The Beta distribution is one of the most commonly used model for rates and proportions. We can assume that the unemployment rate r_{jt} follows a Beta distribution and using the parameterization proposed by Ferrari and Cribari-Neto (2004), we denote by

$$r_{jt}|\mu_{jt},\phi \sim \mathit{Beta}(\mu_{jt},\phi)\,, \qquad j=1,...,28\,, \quad t=1,...,12\,.$$

The probability mass function is given by

$$p(r_{jt}|\mu_{jt},\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu_{jt}\phi)\,\Gamma\big((1-\mu_{jt})\phi\big)}\; r_{jt}^{\mu_{jt}\,\phi-1}(1-r_{jt})^{(1-\mu_{jt})\phi-1}\,, \qquad 0<r_{jt}<1\,,$$

where $0<\mu_{jt}<1$ and $\phi>0$.

In this case, there are several possible choices for the link function, but the most common is the logit function $h(\mu_{jt}) = \log(r_{jt}/(1-r_{jt}))$.

3.5. Multinomial model

The Multinomial logistic regression model is an extension of the Binomial logistic regression model and is used when the variable of interest is multi-category. In this case, it may interest us to model the three categories of the labor market (employment, unemployment and inactivity), giving us the unemployment rate which can be expressed by the ratio between the unemployed and the active people (the sum of the unemployed and employed).

One advantage of the Multinomial model in this problem is the consistency obtained between the three categories of the labor market. The estimated total employment, unemployment and inactivity coincides with the total population. In addition, the same model provides estimates for the rate of employment, unemployment and inactivity.

Assuming that $\boldsymbol{y_{jt}} = (y_{jt1}, y_{jt2}, y_{jt3})$ is the vector of the total in the three categories of the labor market, the Multinomial model can be written as

$$\boldsymbol{y_{jt}}|\,n_{jt},\boldsymbol{P_{jt}} \sim \mathit{Multinomial}(n_{jt},\boldsymbol{P_{jt}})\,, \qquad j=1,...,28\,, \quad t=1,...,12\,,$$

where n_{jt} is the number of individuals in the area j and quarter t, and $\boldsymbol{P_{jt}} = (P_{jt1}, P_{jt2}, P_{jt3})$ is the vector of proportions of employed, unemployed and inactive, where $P_{jt3} = 1-(P_{jt1}+P_{jt2})$.

The probability mass function is given by

$$p\big(y_{jt1}, y_{jt2}, y_{jt3}\,|\,n_{jt}, P_{jt}\big) \;=\; \frac{n_{jt}!}{y_{jt1}!\,y_{jt2}!\,y_{jt3}!}\; P_{jt1}^{y_{jt1}}\,P_{jt2}^{y_{jt2}}\,P_{jt3}^{y_{jt3}}\,,$$

where

$$y_{jtq}\in\mathbb{N}: \;\; \sum_{q}^{3} y_{jtq} = n_{jt}\,, \quad q=1,2,3\,.$$

The most common link function is the log of P_{jtq}, defined as $\eta_{jtq} = \log(P_{jtq}/P_{jt3})$, $q=1,2$.

4. APPLICATION TO THE PORTUGUESE LABOR FORCE SURVEY DATA

4.1. Results

This section provides the results of applying five models for the estimation of the total unemployed and unemployment rate to the NUTS III regions of Portugal.

The Poisson, Binomial, Negative Binomial and Beta models were implemented using the R package *R-INLA*, while the Multinomial model was implemented based on MCMC methods using the R package *R2OpenBUGS*.

When the Multinomial regression model was combined with the predictor given in (1), some convergence problems arose, due to its complexity. For this reason, the effects w_{jt} and ϵ_{jt} were replaced by the unstructured area and time effects u_j and v_t, where it was assumed

$$u_j \sim N(0, \sigma_u^2)\,,$$
$$v_t \sim N(0, \sigma_v^2)\,,$$

with the following prior information

$$\log \tau_u \sim \log Gamma(1, 0.0005)\,,$$
$$\log \tau_v \sim \log Gamma(1, 0.0005)\,.$$

Due to the differences in the model structure and the computational methods used for the Multinomial model, the comparative analysis of results for this model should be done with some extra care.

The posterior mean of the parameters and hyperparameters of each model as well as the standard deviation and the quantile 2.5% and 97.5% are presented in Tables 1, 2, 3, 4 and 5. We can see that the covariates *GDP* and *secondary sector* are not significant for any of the models applied. However, the value obtained for *Deviance Information Criterion* (DIC) increases considerably without the inclusion of these variables, so we decided to include them.

We observe that the IEFP is significant in all of the models applied, as expected. The number of enterprises per 100 000 inhabitants has a negative effect on the increase of unemployment. The population structure has also a significant effect. The proportion of individuals that are female and aged between 24 and 34 years has a positive effect on the increase of unemployment. On the other hand, the proportion of individuals that are female and over 49 years has a negative

effect. These tendencies are probably due to young unemployment in the first case and to the fact that the age group +49 includes most of the inactive people, in the second case.

Table 1: Posterior mean, standard deviation and 95% credibility interval for the parameters and hyperparameters of Poisson model.

Poisson				
Parameter	**Mean**	**SD**	**2.5Q**	**97.5Q**
(Intercept)	−2.83	0.01	−2.85	−2.81
Companies	−0.01	0.02	−0.05	0.02
Primary sector	−0.02	0.72	−1.45	1.40
Secondary sector	0.02	0.21	−0.39	0.43
GDP	0.00	0.00	0.00	0.00
IEFP	10.05	0.96	8.17	11.93
SA6	4.30	1.34	1.65	6.93
SA8	−1.55	0.57	−2.65	−0.42
τ				
τ_{w_2}	25047.76	20819.39	3297.22	79744.21
τ_{w_1}	25.77	9.52	11.91	48.78
τ_ϵ	22082.79	19692.19	2213.88	73957.91

Table 2: Posterior mean, standard deviation and 95% credibility interval for the parameters and hyperparameters of Negative Binomial model.

Negative Binomial				
Parameter	**Mean**	**SD**	**2.5Q**	**97.5Q**
(Intercept)	−2.83	0.01	−2.86	−2.81
Companies	−0.01	0.02	−0.05	0.02
Primary sector	0.13	0.73	−1.33	1.57
Secondary sector	−0.04	0.23	−0.48	0.41
GDP	0.00	0.00	0.00	0.00
IEFP	10.20	1.48	7.28	13.09
SA6	3.97	2.01	0.02	7.91
SA8	−2.11	0.73	−3.54	−0.67
τ	48.57	5.44	38.69	60.05
τ_{w_2}	22946.14	20085.32	2453.14	75579.11
τ_{w_1}	32.77	14.26	13.08	68.00
τ_ϵ	22641.11	19924.30	2405.53	74957.89

All the considered models give very good fit to the data and their temporal predictions are also satisfactory. Here we report on several model fitting aspects of the Binomial model. Similar results for the other models are given in the Supplementary Material.

Table 3: Posterior mean, standard deviation and 95% credibility interval for the parameters and hyperparameters of Binomial model.

Binomial				
Parameter	**Mean**	**SD**	**2.5Q**	**97.5Q**
(Intercept)	−1.97	0.01	−2.00	−1.95
Companies	−0.04	0.02	−0.07	0.00
Primary sector	0.54	1.01	−1.47	2.52
Secondary sector	−0.11	0.28	−0.67	0.45
GDP	0.00	0.00	0.00	0.00
IEFP	12.63	1.11	10.47	14.81
SA6	4.38	1.47	1.50	7.26
SA8	−1.11	0.64	−2.37	0.16
τ				
τ_{w_2}	20736.79	19243.68	2070.46	71553.26
τ_{w_1}	11.77	3.90	5.70	20.85
τ_ϵ	19143.06	18555.71	1460.60	68268.97

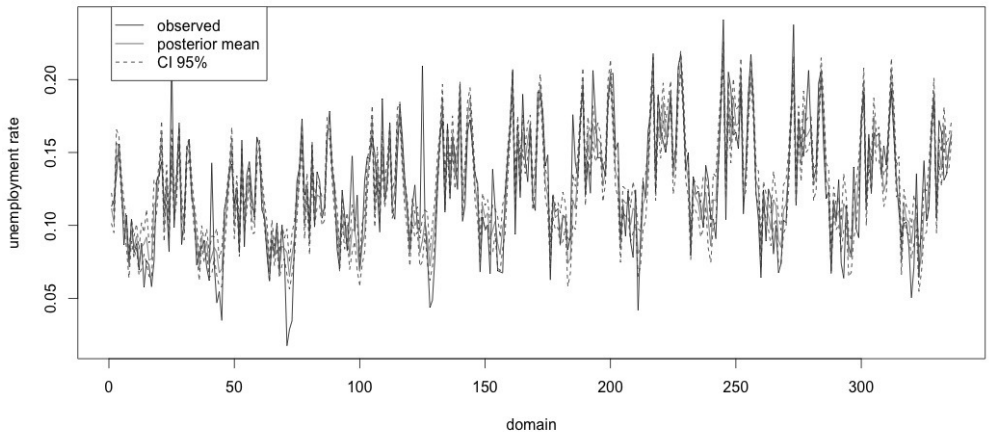
Table 4: Posterior mean, standard deviation and 95% credibility interval for the parameters and hyperparameters of Beta model.

Beta				
Parameter	**Mean**	**SD**	**2.5Q**	**97.5Q**
(Intercept)	−1.98	0.01	−2.00	−1.95
Companies	−0.03	0.03	−0.08	0.02
Primary sector	0.69	1.09	−1.50	2.83
Secondary sector	−0.20	0.31	−0.82	0.42
GDP	0.00	0.00	0.00	0.00
IEFP	12.22	1.82	8.64	15.78
SA6	0.85	1.84	−2.77	4.47
SA8	−2.37	0.68	−3.70	−1.03
τ	206.61	16.80	174.43	240.37
τ_{w_2}	20012.58	19075.29	1750.73	70491.89
τ_{w_1}	11.33	4.61	5.40	23.02
τ_ϵ	20497.48	19404.83	1715.97	71768.00

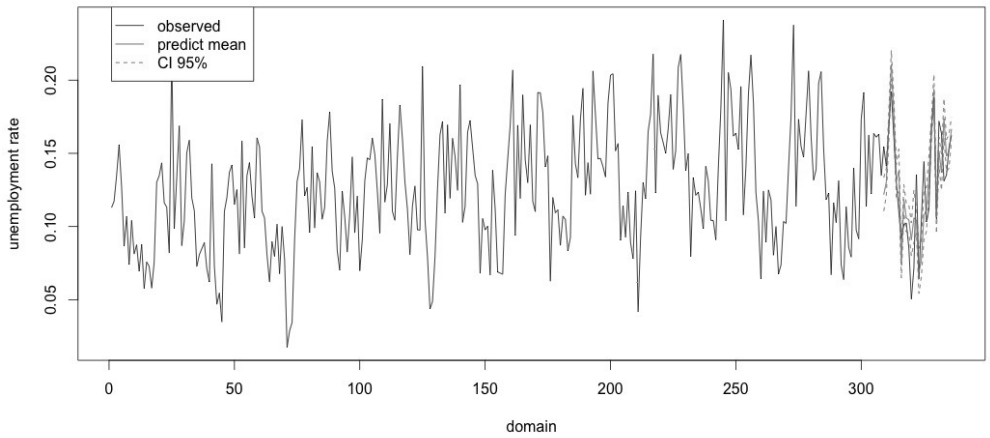
Table 5: Posterior mean, standard deviation and 95% credibility interval for the parameters and hyperparameters of Multinomial model.

Multinomial				
Parameter	**Mean**	**SD**	**2.5Q**	**97.5Q**
(Intercept)	−1.74	0.26	−2.16	−1.20
Companies	0.01	0.02	−0.04	0.05
Primary sector	3.99	5.38	−1.04	14.94
Secondary sector	−0.77	0.77	−2.27	0.19
GDP	0.00	0.00	0.00	0.00
IEFP	8.93	1.59	6.02	12.00
SA6	4.77	1.44	1.97	7.58
SA8	−2.38	0.64	−3.74	−1.22
τ_v	2206.25	2979.60	3.32	9519.00
τ_u	33.57	24.87	1.77	78.11

Figure 3 (a) gives the observed and adjusted values from the Binomial model together with their 95% credible intervals, whereas Figure 3 (b) gives the predictions to the 4th quarter of 2013 together with their 95% credible intervals. We see that the adjusted values are very close to the observed ones. The domains are sorted at first by quarter and then by region. This is the reason for the identical behavior in each 28 domains (corresponding to the NUTS III regions). The graphs show a slight increase on the unemployment rate until the 1st quarter of 2013 and then a decrease until the 4th quarter of 2013.



(a)



(b)

Figure 3: (a) Observed and adjusted values (mean and 95% CI) of unemployment rate for the 336 domains (336 = 28 NUTS III × 12 quarters); (b) Observed and predicted values (the posterior mean and 95% CI) of the unemployment rate. The prediction is made for the 4th quarter of 2013 which is highlighted red.

The map of the Figure 4 allows for a better understanding of the regional difference between the observed and fitted values.

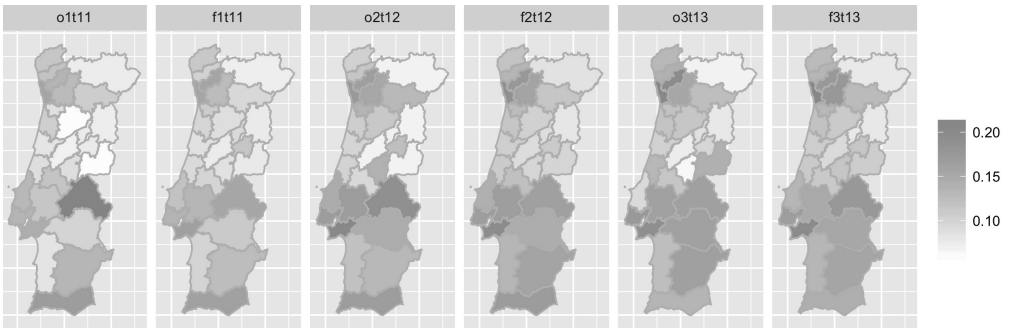


Figure 4: Maps of observed and fitted values of the unemployment rate for the 1st quarter of 2011, 2nd quarter of 2012 and 3rd quarter of 2013.

4.2. Diagnosis

Some predictive measures can be used for an informal diagnostic, such as *Conditional Predictive ordinates* (CPO) and Probability Integral Transforms (PIT; Gelman *et al.*, 2004). Measure CPO_i is defined as $\pi(y_i|y_{-i})$ where y_{-i} is the vector y without observation y_i, while the measures PIT_i are obtained by $Prob(y_i^{new} \leq y_i|y_{-i})$. Unusually large or small values of this measure indicate possible outliers. Moreover, a histogram of the PIT value which is very different from the uniform distribution indicates that the model is questionable.

The implementation of these measures in an MCMC approach is very heavy and requires a high computational time. For this reason, we present only results for the models implemented with the INLA.

Figure 5 shows the graphs of the PIT values versus domain ($28 \times 12 = 336$) and the histogram of the PIT values for Poisson, Binomial, Negative Binomial and Beta models. We see that the histogram for the PIT values based on the Poisson and Binomial models presents a fairly uniform behavior, but this is not the case with the Negative Binomial and Beta distributions. This suggests that these last two models may not be suitable for data in analysis.

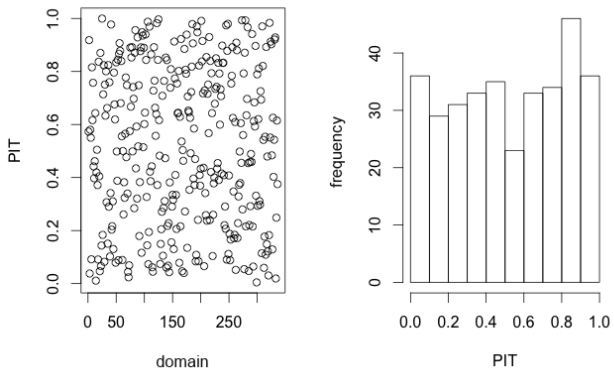


Figure 5: Graphs of the PIT values versus domain ($28 \times 12 = 336$) and the histogram of the PIT values.

The predictive quality of the models can be performed using a cross-validated logarithmic score given by the symmetric of the mean of the logarithm of CPO values (Martino and Rue, 2010). High CPO values indicate a better quality of prediction of the respective model. The logarithmic of the CPO values are given in table 6. Accordingly, the Beta model has the least predictive quality.

The diagnosis of the Multinomial model was based on graphical visualization and on *Potential Scale Reduction Factor* (Brooks and Gelman, 1997). No convergence problems were detected.

Table 6: Logarithmic score.

Model	**log score**
Poisson	3.33
Negative Binomial	3.51
Binomial	3.34
Beta	−2.39

4.3. Comparison

In order to compare the studied models, we use the *Deviance Information Criterion* (DIC) proposed by Spiegehalter *et al.* (2002). This is a criterion which aims to achieve a balance between the adequacy of a model and its complexity. It is defined by $DIC = \bar{D} + p_D$ where $\bar{D}$ is the posterior mean deviance of the model and p_D is the effective number of parameters. The model with the smallest

value of DIC is the one with a better balance between the model adjustment and complexity.

The values of DIC, p_D and $\bar{D}$ are presented in table 7. The Multinomial model features a higher DIC value, but it should be noted that this model requires an adjustment of the total of employed, unemployed and inactive people, unlike the other models.

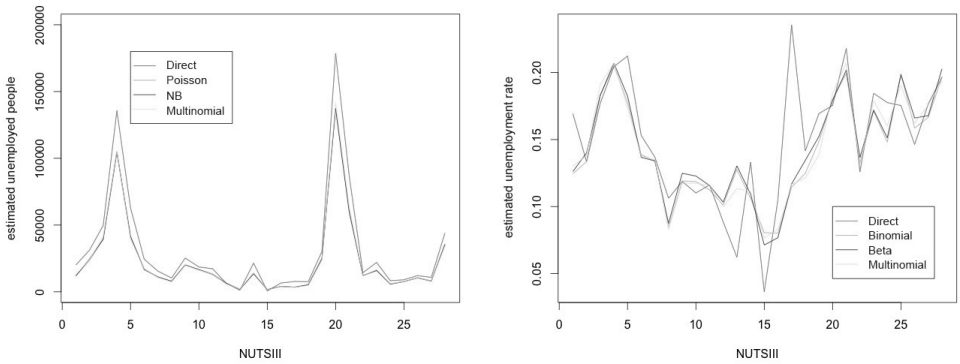
Among the models used for modeling of total, the Poisson model is the one with the lower value of DIC, which would suggest that it should be preferable to the Negative Binomial model. However, Geedipally *et al.* (2013) explains that the value of this measure is affected by the parameterization of the model, which may influence the values obtained by the Negative Binomial and Beta models, since the software used permits different parameterizations in these cases.

Table 7: DIC, effective number of parameters, and posterior mean of the deviance.

Model	DIC	p_D	$\bar{D}$
Poisson	2240.4	30.4	2210.0
Negative Binomial	2374.9	25.4	2349.5
Binomial	2241.4	32.5	2208.9
Beta	−1607.4	31.1	−1638.4
Multinomial	4976.0	81.5	4894.5

Figure 6 shows that the Poisson, Negative Binomial and Multinomial models produced very similar estimates for the total unemployed, while Binomial, Beta and Multinomial models produced similar estimates for unemployment rate. We can also note that these estimates are smoother than the estimates obtained by the direct method. This property is prominently displayed in the estimation of the unemployment rate, and justifies the fact that the estimates of the total produced by the models are lower than the estimates produced by the direct method for large values of unemployment, and higher for small values (regions 13 and 15).

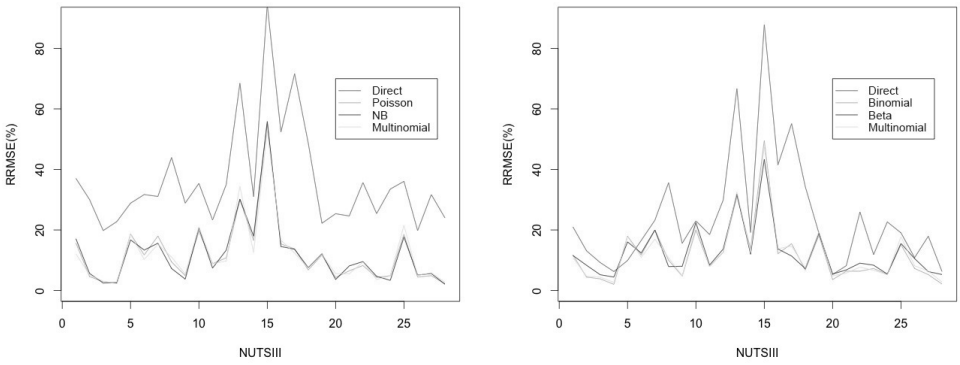
Regions 4 and 20, which correspond to Grande Porto and Grande Lisboa, are those with the highest population size. This fact explains the high values of unemployment totals. On the other hand, regions 13 and 15, which correspond to Pinhal Interior Sul and Serra da Estrela, are those with the lowest population size. It is interesting to observe that the regions with the greatest difference between the unemployment rate estimated by the direct method and the rate estimated by the studied models are those with the lowest population sizes (Pinhal Interior Sul, Serra da Estrela and Beira Interior Sul), which are represented in the graph by the numbers 13, 15 and 17.



(a) Totals (b) Rates

Figure 6: Estimates for the total unemployed (through Poisson, Negative Binomial, Multinomial, and direct method) and the unemployment rate (through Binomial, Beta, Multinomial, and direct method).

The Relative Root Mean Square Error (RRMSE) allows for a comparison of the models studied and the direct method. A lower value of RRMSE indicates a better balance between variability and bias. The graph of Figure 7 reveals a wide discrepancy between the direct method and the applied models.



(a) Totals (b) Rates

Figure 7: RRMSE estimates for the total unemployed and the unemployment rate.

Note that, for most models, the NUTS III region 15, which corresponds to Serra da Estrela (see Appendix), presents the highest value RRMSE. This result can be explained in part by the reduced population size of the region. The opposite is true for regions with high dimensional population such as Porto (Region 4), Grande Lisboa (region 20) and Algarve (region 28).

The high values of RRMSE of unemployment rate estimates by the direct method for regions 13, 15 and 17, can explain the big differences found between the methods in these regions (Figure 7 (b)). These results reinforce the idea that the direct method is inadequate for the estimation in small areas. On the other hand, the models studied show a maximum value of RRMSE of the unemployment rate estimates that corresponds to almost half of the value obtained by the direct method.

5. DISCUSSION

We studied the application of five spatio-temporal models within a Bayesian approach for the estimation of both the total and the rate of unemployment of NUTS III regions. We realized that one of the features of model based methods is the smoothing of the variation across time and space. This feature brings these models closer to reality.

The estimates obtained by these models were reasonable when compared with the direct method, which presented higher values of RRMSE.

Models under study presented much lower values of RRMSE than the direct method for regions with a small population size. This feature shows that these models can be a good alternative to small area estimation and in particular for the NUTS III regions of Portugal.

The Negative Binomial and Beta models presented diagnostic problems in the analysis of empirical distribution of the PIT. A non uniform distribution of the PIT revealed that the predictive distribution is not coherent with the data.

Among the models under study, the Multinomial model seems to be the most suitable for this problem. The estimates obtained for the unemployment totals are similar to those obtained by the other models, but they produce estimates for the total of employed as well as inactive people simultaneously, in a way that is consistent with the population estimates. In this way, we can directly obtain the estimates of the employment and unemployment rates.

APPENDIX

Table 8: NUTS II and NUTS III regions of Continental Portugal.

NUTS III region			NUTS II region
Index	Code	Designation	Designation
1	111	Minho-Lima	Norte
2	112	Cávado	
3	113	Ave	
4	114	Grande Porto	
5	115	Tâmega	
6	116	Entre Douro e Vouga	
7	117	Douro	
8	118	Alto Trás-os-Montes	
9	161	Baixo Vouga	Centro
10	162	Baixo Mondego	
11	163	Pinhal Litoral	
12	164	Pinhal Interior Norte	
13	166	Pinhal Interior Sul	
14	165	Dão-Lafões	
15	167	Serra da Estrela	
16	168	Beira Interior Norte	
17	169	Beira Interior Sul	
18	16A	Cova da Beira	
19	16B	Oeste	
20	171	Grande Lisboa	Lisboa
21	172	Península de Setúbal	
22	16C	Médio Tejo	Centro
23	185	Lezíria do Tejo	Alentejo
24	181	Alentejo Litoral	
25	182	Alto Alentejo	
26	183	Alentejo Central	
27	184	Baixo Alentejo	
28	150	Algarve	Algarve

ACKNOWLEDGMENTS

This work was supported by the project UID/MAT/00006/2013 and the PhD scholarship SFRH/BD/92728/2013 from Fundação para a Ciência e Tecnologia. Instituto Nacional de Estatística and Centro de Estatística e Aplicações da Universidade de Lisboa are the reception institutions. We would like to thank professor Antónia Turkman for her help.

NOTE

This study is the responsibility of the authors and does not reflect the official opinions of Instituto Nacional de Estatística.

REFERENCES

[1] Besag, J.; York, J. and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics, *Annals of the Institute of Statistical Mathematics*, **43**, 1–59.

[2] Best, N.; Richardson, S. and Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping, *Statistical Methods in Medical Research*, **14**, 35–59.

[3] Blangiardo, M.; Cameletti, M.; Baio, G. and Rue, H. (2013). Spatial and spatio-temporal models with R-INLA, *Spatial and Spatio-temporal Epidemiology*, **7**, 39–55.

[4] Brooks, S.P. and Gelman, A. (1997). General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, **7**, 434–455.

[5] Chambers, R.; Salvati, N. and Tzavidis, N. (2016). Semiparametric small area estimation for binary outcomes with application to unemployment estimation for local authorities in the UK, *Journal of the Royal Statistical Society, Series A*, **179**, 453–479.

[6] Choundry, G.H. and Rao, J.N.K. (1989). Small area estimation using models that combine time series and cross sectional data, *Journal of Statistics Canada Symposium on Analysis of Data in Time*, 67–74.

[7] Datta, G.S. and Ghosh, M. (1991). Bayesian prediction in linear models: application to small area estimation, *Annals of Statistics*, **19**, 1748–1770.

[8] Ferrari, S. and Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions, *Journal of Applied Statistics*, **31**, 799–815.

[9] Geedipally, S.R.; Lord, D. and Dhavala, S.S. (2014). A caution about using deviance information criterion while modeling traffic crashes, *Safety Science*, **62**, 495–498.

[10] Gelman, A.; Carlin, J.; Stern, H. and Rubin, D. (2004). *Bayesian Data Analysis*, Second Edition, London, Chapman and Hall.

[11] Jiang, J. and Lahiri, P. (2006a). Estimation of finite population domain means: A model-assisted empirical best prediction approach, *Journal of the American Statistical Association*, **101**, 301–311.

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

[13] Lopez-Vizcaino, E.; Lombardia, M.J. and Morales, D. (2015). Small area estimation of labour force indicators under a multinomial model with correlated time and area effects, *Journal of the Royal Statistical Society, Series A*.

[14] Martino S. and Rue, H. (2010). Case Studies in Bayesian Computation using INLA, *Complex data modeling and computationally intensive statistical methods* (R-code).

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

[16] Rao, J.N.K. (2003). *Small area estimation*, Wiley.

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

[18] Singh, B.; Shukla, G. and Kundu, D. (2005). Spatio-temporal models in small area estimation, *Survey Methodology*, **31**, 183–195.

[19] Spiegelhalter, D.; Best, N.; Bradley, P. and van der Linde, A. (2002). Bayesian measure of model complexity and fit (with discussion), *Journal of the Royal Statistical Society*, Series B, **64**(4), 583–639.

REVSTAT – STATISTICAL JOURNAL

Background

Statistics Portugal (INE, I.P.), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT - STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another external.

— The only working language allowed will be English. — Four volumes are scheduled for publication, one in January, one in April, one in July and the other in October.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract and Indexing Services

The REVSTAT is covered by the following abstracting/indexing services:

- Current Index to Statistics
- Google Scholar
- Mathematical Reviews
- Science Citation Index Expanded
- Zentralblatt für Mathematic

Instructions to Authors, special-issue editors and publishers

The articles should be written in English and may be submitted in two different ways:

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt) and to one of the two Editors or Associate Editors, whose opinion the author wants to be taken into account, together to the following e-mail address: revstat@fc.ul.pt

- By sending the paper in PDF format to the Executive Editor (revstat@ine.pt), together with the corresponding PDF or PostScript file to the following e-mail address: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Manuscripts (text, tables and figures) should be typed only in black on one side, in double-spacing, with a left margin of at least 3 cm and with less than 30 pages. The first page should include the name, institution and address of the author(s) and a summary of less than one hundred words, followed by a maximum of six key words and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style. This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to PC Windows System (Zip format), Macintosh, Linux and Solaris Systems (StuffIt format), and Mackintosh System (BinHex Format), are available in the REVSTAT link of the Statistics Portugal Website: http://www.ine.pt/revstat/inicio.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: revstat@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Editorial Board