Brazilian Journal of Probability and Statistics
2016, Vol. 30, No. 3, 385–400
DOI: 10.1214/15-BJPS285

Bias-corrected maximum likelihood estimation of the parameters of the complex Bingham distribution

Luiz H. G. Dore[a], Getulio J. A. Amaral[b], Jorge T. M. Cruz[b] and Andrew T. A. Wood[c]

[a]*Universidade Federal de Sergipe*
[b]*Universidade Federal de Pernambuco*
[c]*University of Nottingham*

Abstract. In this paper, some bias correction methods are considered for parameter estimation of the complex Bingham distribution. The first method relies on the bias correction formula proposed by Cordeiro and Klein [*Statistics & Probability Letters* **19** (1994) 169–176]. The second method uses the formulas proposed by Kume and Wood [*Statistics & Probability Letters* **77** (2007) 832–837] for calculating the derivatives of the log likelihood function. The third method is based on the saddlepoint approximation proposed by Kume and Wood [*Biometrika* **92** (2005) 465–476]. Bootstrap bias correction methods due to Efron [*The Annals of Statistics* **7** (1979) 1–26] are also considered. Simulation experiments are used to compare the bias correction methods. In all cases, the analytical and bootstrap bias correction methods have smaller mean square errors. Since the dominant eigenvalue is used to obtain the mean shape, which has practical relevance, it is a key issue for comparing the estimators. The numerical results indicate that the bootstrap methods have a slightly better performance for the dominant eigenvalue.

1 Introduction

The Bingham distribution was introduced by Bingham (1974) as a generalization of the Dimroth–Watson distribution (Watson, 1965). Let x be a point on the unit sphere and $f(x)$ its probability density function. The Bingham distribution has antipodal symmetry, which means that $f(x) = f(-x)$ for all unit vectors x. It makes the Bingham distribution suitable for modelling axes on the sphere where the vectors x and $-x$ represent the same axis. The Bingham distribution can also be equivalently defined as follows: if x is a random vector from a multivariate normal distribution with zero mean and conditioned to have unit length, one can say that the distribution of x is the Bingham distribution (Mardia and Jupp, 2000).

A complex version of the Bingham distribution was presented by Kent, Constable and Er (2004) for use in statistical shape analysis for objects in 2 dimensions. In this work, Kent, Constable and Er (2004) derived several properties of the

Key words and phrases. Pre-shapes, non-Euclidean spaces, Cox and Snell formula.
Received February 2014; accepted February 2015.

complex Bingham distribution, including invariance under scalar rotation, meaning that $f(e^{i\theta}z) = f(z)$, where f is the density function of the complex Bingham distribution. They also developed likelihood estimation and provided other relevant results. Simulation of the complex Bingham distribution is covered by Kent, Constable and Er (2004).

Some specific details should be considered for the complex Bingham distribution. Kume and Wood (2005) obtained a mathematical expression for the constant of the complex Bingham distribution. Kume and Wood (2007) obtained the derivatives of the normalizing constant of the Bingham distribution. These results are used to obtain bias corrected estimators.

Amaral, Florez and Cysneiros (2013) introduced an influence measure in statistical shape analysis related to the complex Bingham distribution. This influence measure is based on the Cook's distance and the complex Bingham model is assumed for the sample. Some results from that work are used in this paper.

There is motivation for bias correction in statistical shape analysis. In this context, small sample sizes happens frequently (Dryden and Mardia, 1998). When the sample size is small, it is relevant to study the behaviour of maximum likelihood estimators (MLEs). Bias correction is one of the most important topics in those studies. Cox and Snell (1968) have derived a general formula for analytical bias correction. Cordeiro and Klein (1994) have proposed one simplified version of that formula. The Cordeiro and Klein (1994) formula is used to bias correct the MLEs of the complex Bingham distribution.

The rest of the paper is organized as follows. Some relevant details about the complex Bingham distribution are reviewed in Section 2. In Section 3, the Cordeiro and Klein (1994) formula is applied to the complex Bingham distribution. The Kume and Wood (2007) formula is used in Section 4. The saddlepoint approximation is considered in Section 5. Bootstrap bias correction is defined in Section 6. Some numerical results are presented in Section 7 and some conclusions are given in Section 8.

2 Some properties of the complex Bingham distribution

It is relevant to motivate how the complex Bingham distribution arises in practical situations. A brief discussion is given in this section, and further details can be found in Dryden and Mardia (1998), Small (1996) and Kent (1994).

Let

$$Y = \begin{pmatrix} y_{1,1} & y_{1,2} \\ \vdots & \vdots \\ y_{k,1} & y_{k,2} \end{pmatrix} \tag{2.1}$$

be the representation of k landmarks of an object in two dimensions.

In shape analysis, it is necessary to remove the effects of translation, scale and rotation.

The translation and scale effects are often removed. However, the rotation effect is removed only when graphical representation is necessary.

To simplify the mathematical operations, the first step that one should perform is to transform the matrix (2.1) in a complex vector. It is given by

$$w^0 = \begin{pmatrix} y_{1,1} + i y_{1,2} \\ \vdots \\ y_{k,1} + i y_{k,2} \end{pmatrix}. \tag{2.2}$$

The effect of translation is removed by multiplying the vector (2.2) by the Helmert sub matrix H which is defined as follows. The matrix H is a $(k-1) \times k$ matrix which the j row is given by

$$(h_j, \dots, h_j, -j h_j, 0, \dots, 0), \qquad h_j = -\{j(j+1)\}^{-1/2},$$

with $j = 1, \dots, k-1$, where the number of zeros elements in the row j is equal to $k - j - 1$.

So the translation and scale effects are removed by the operation

$$z = \frac{H w^0}{|H w^0|}.$$

The vector z is called a pre-shape, and the complex Bingham distribution is often used as a model for a random sample of pre-shapes $z_1, \dots, z_n$.

The probability density function (p.d.f.) of a complex Bingham distribution is given by

$$f(z) = c(A)^{-1} \exp(z^\star A z), \qquad z \in \mathbb{C}S^{k-1}, \tag{2.3}$$

where z is a complex unit vector, $z^\star = \bar{z}^T$ is the complex conjugate of the transpose of z, A is a $(k-1) \times (k-1)$ Hermitian matrix and $c(A)$ is a normalizing constant. The matrix A is a parameter matrix and $\operatorname{diag}(\lambda_1, \dots, \lambda_{k-1})$ will denote its eigenvalues.

Let $z_1, \dots, z_n$ be a random sample from the density (2.3). After some algebraic simplifications (Dryden and Mardia, 1998), the likelihood of the p.d.f. (2.3) is given by

$$l(\lambda) = \sum_{r=1}^{k-1} l_r \lambda_r - n \log c(\Lambda), \tag{2.4}$$

where $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_{k-1})$, $l_1, \dots, l_{k-1}$ are the eigenvalues of $\hat{S} = \sum_{i=1}^{n} z_i z_i^\star$, $c(\Lambda) = 2\pi^{k-1} \sum_{r=1}^{k-1} a_r e^{\lambda_r}$ and $a_r^{-1} = \prod_{j \neq r} (\lambda_r - \lambda_j)$.

One important property of the complex Bingham distribution is that the matrices A and $A + kI$ define the same distribution.

Kent, Constable and Er (2004) explains how to generate random vectors from the density (2.3). We use the first algorithm proposed by Kent, Constable and Er (2004), which is described as follows:

(1) Generate $(k-1)$ uniform $(0, 1)$ random numbers: $U_1, \ldots, U_{k-1}$.
(2) Compute $S_i = -(1/\lambda_j)\log(1 - U_j(1 - e^{-\lambda_j}))$, $S' = (S_1, \ldots, S_{k-1})$, noting that the S_i are independent truncated exponential.
(3) If $\sum_{j=1}^{k-1} S_j < 1$, set $S = S'$. Otherwise, reject S' and go to 1.
(4) Generate independent angles $\theta_j \sim U[0, 2\pi)$, $j = 1, \ldots, k-1$.
(5) Calculate $z_j = s_j^{1/2}\exp(i\theta_j)$, $j = 1, \ldots, (k-1)$.

The final vector $(z_1, \ldots, z_{k-1})$ comes from the density (2.3).

3 Analytical bias correction

Bias correction is a well explored topic. Cordeiro and Cribari-Neto (2014) present an account about this theme. The motivation for this paper and other relevant information come from that work.

One important issue is that maximum likelihood estimates (MLE) are often biased. This error is systematic (Cordeiro and Cribari-Neto, 2014). Since the bias order of the MLE is n^{-1}, it is a problem when the sample size is small. A bias correction formula for $O(n^{-1})$ bias has been introduced by Bartlett (1953) for one-parameter models. Cox and Snell (1968) proposed a bias correction formula for multi-parameter models.

Since most of the data set considered by Dryden and Mardia (1998) have small sample sizes, there is a motivation for bias correction in complex Bingham models.

Let $z_1, \ldots, z_n$ a random sample from the density (2.3), where $\lambda^T = (\lambda_1, \ldots, \lambda_{k-1})$ is the parameter vector which is defined in (2.4).

According to Cordeiro and Klein (1994), the second-order bias correction of the parameter vector λ is given by

$$B(\hat{\lambda}_r) = \sum_{s,t,u} \kappa^{ru}\kappa^{st}\left(\kappa_{rs}^{(t)} - \frac{1}{2}\kappa_{rst}\right),$$

where:

(1) $\hat{\lambda}_r$ is the maximum likelihood estimate (MLE) estimate for the rth component of λ;
(2) $-\kappa^{ru}$ is the element (r, u) of the inverse of the expected Fisher information matrix;
(3) $\kappa_{rs}^{(t)} = \frac{\partial\kappa_{rs}}{\partial\lambda_t}$, where $-\kappa_{rs}$ is the (r, s) element of the Fisher information matrix;
(4) $\kappa_{rst} = E(\frac{\partial^3 l}{\partial\lambda_r\,\partial\lambda_s\,\partial\lambda_t})$.

The second partial derivative of (2.4) does not depend on the sample. Therefore,

$$\kappa_{rs} = \frac{\partial^2 l}{\partial \lambda_r \, \partial \lambda_s}$$

and

$$\kappa_{rst} = \frac{\partial^3 l}{\partial \lambda_r \, \partial \lambda_s \, \partial \lambda_t} = \frac{\partial}{\partial \lambda_t}\left(\frac{\partial^2 l}{\partial \lambda_r \, \partial \lambda_s}\right) = \frac{\partial \kappa_{rs}}{\partial \lambda_t} = \kappa_{rs}^{(t)}.$$

Thus, in this case , the second-order bias correction is given by

$$B(\hat{\lambda}_r) = \frac{1}{2}\sum_{s,t,u} \kappa^{ru}\kappa^{st}\kappa_{rst}. \tag{3.1}$$

To obtain the terms of (3.1) it is necessary to use some quantities from Amaral, Florez and Cysneiros (2013). So κ_{rs} is given by

$$\kappa_{rs} = \frac{n}{c^2(\Lambda)}\left[c_r(\Lambda)c_s(\Lambda) - c_{rs}(\Lambda)\right],$$

where $c_r(\Lambda) = \frac{\partial c}{\partial \lambda_r}$ and $c_{rs}(\Lambda) = \frac{\partial^2 c}{\partial \lambda_r \, \partial \lambda_s}$. Set $c_{rst}(\Lambda) = \frac{\partial^3 c}{\partial \lambda_r \, \partial \lambda_s \, \partial \lambda_t}$, we have

$$\kappa_{rst} = \frac{n}{c^2(\Lambda)}\Bigg\{\left[c_{rt}(\Lambda)c_s(\Lambda) + c_{st}(\Lambda)c_r(\Lambda) - c_{rst}(\Lambda)\right] - 2\frac{c_t(\Lambda)}{c(\Lambda)}\left[c_r(\Lambda)c_s(\Lambda) - c_{rs}(\Lambda)\right]\Bigg\}.$$

The terms of equation (3.1) are computed in the Appendix.

4 Formulas proposed by Kume and Wood (2007) for the derivatives

To obtain the derivatives of the complex Bingham normalizing constant is not an easy task. Kume and Wood (2007) have shown that the derivative of that constant is proportional to the normalizing constant of a real Bingham distribution. So they use sanddlepoint approximations to compute the derivatives of the complex Bingham normalizing constant.

The results of Kume and Wood (2007) can be used to obtain the derivatives of the likelihood as follows. The formula of the derivatives is given by

$$\frac{\partial^n}{\partial \lambda_1^{m_1}, \ldots, \partial_{\lambda_p}^{m_p}}\big(c(\Lambda)\big) = (-1)^n 2\pi^p \sum_{i=1}^{p} (-1)_i^m \exp(-\lambda_i) S(m_i),$$

where

$$S(m_i) = \sum_{J_{0(i)} \geq 0, |J_{0(i)}| = m_i} \frac{m_i!}{j_0!, \ldots, j_{i-1}!, \ldots, j_p!}(-1)^{j_0} \prod_{r \neq i} \frac{(m_r + j_r)!}{(\lambda_r - \lambda_i)^{1+m_r+j_r}},$$

where $j_0(i) = (j_0, j_1, \ldots, j_{i-1}, j_{i+1}, \ldots, j_p)$ with $j_k > 0, \forall k$, and $|J_{0(i)}|$ denoting the sum of the components of $J_{0(i)}$.

For each i, $i = 1, \ldots, p$, it must be considered all the combinations of the values $j_0, \ldots, j_p$ such that $|J_{0(i)}| = m_i$, where m_i is the number of times that $c(\Lambda)$ is differentiated by λ_i. For example, for the computation of $\frac{\partial^2 c(\Lambda)}{\partial \lambda_t^2}$, we have $m_t = 2$, $m_i = 0\ \forall i \neq t$ and $n = 2$. Hence, for each $i \neq t$, there is only one combination $J_{0(i)}$ (all components are equal to zero) and the value of $S(m_i)$ is the following:

$$S(m_i) = S(0) = \frac{-2a_i}{(\lambda_t - \lambda_i)^2}. \tag{4.1}$$

If $i = t$, we may enumerate four types of combinations:

(1) $j_0 = 2$;
(2) $j_0 = 0$ and $j_k = 2$, for some $k > 0$;
(3) $j_0 = 1$ and $j_k = 1$ for some $k > 0$;
(4) $j_{k_1} = j_{k_2} = 1$ for some $k_1, k_2 > 0$ with $k_1 \neq k_2$.

(1) There is only one combination: $J_{0(t)} = (2, 0, \ldots, 0)$. Thus,

$$\frac{m_t!}{j_0!, \ldots, j_{i-1}!, \ldots, j_p!}(-1)^{j_0} \prod_{r \neq t} \frac{(m_r + j_r)!}{(\lambda_r - \lambda_t)^{1+m_r+j_r}} = -a_t. \tag{4.2}$$

(2) There are $p - 1$ combinations: $(0, 2, 0, \ldots, 0), (0, 0, 2, \ldots, 0), \ldots, (0, 0, 0, \ldots, 2)$. For this case,

$$\frac{m_t!}{j_0!, \ldots, j_{i-1}!, \ldots, j_p!}(-1)^{j_0} \prod_{r \neq t} \frac{(m_r + j_r)!}{(\lambda_r - \lambda_t)^{1+m_r+j_r}} = \frac{-2a_t}{(\lambda_k - \lambda_t)^2}. \tag{4.3}$$

The sum of the $p - 1$ combinations is $\sum_{k \neq t} \frac{2a_t}{(\lambda_k - \lambda_t)^2} = -2g_t a_t$.

(3) There are $p - 1$ combinations: $(1, 1, 0, \ldots, 0), (1, 0, 1, \ldots, 0), \ldots, (1, 0, 0, \ldots, 1)$. Therefore,

$$\frac{m_t!}{j_0!, \ldots, j_{i-1}!, \ldots, j_p!}(-1)^{j_0} \prod_{r \neq t} \frac{(m_r + j_r)!}{(\lambda_r - \lambda_i)^{1+m_r+j_r}} = \frac{2a_t}{\lambda_k - \lambda_t}. \tag{4.4}$$

The sum of the $p - 1$ combinations is $\sum_{k \neq t} \frac{2a_t}{\lambda_k - \lambda_t} = -2b_t a_t$.

(4) There are $\frac{(p-1)(p-2)}{2}$ combinations $(0, 1, 1, 0, \ldots, 0, 0), (0, 1, 0, 1, \ldots, 0, 0), \ldots, (0, 0, 0, 0, \ldots, 1, 1)$. So,

$$\begin{aligned} &\frac{m_t!}{j_0!, \ldots, j_{i-1}!, \ldots, j_p!}(-1)^{j_0} \prod_{r \neq t} \frac{(m_r + j_r)!}{(\lambda_r - \lambda_i)^{1+m_r+j_r}} \\ &\quad = -2a_t \frac{1}{\lambda_t - \lambda_{k_1}} \frac{1}{\lambda_t - \lambda_{k_2}}. \end{aligned} \tag{4.5}$$

The sum of this value over all these combinations is

$$-\sum_{k_1\neq t}\sum_{k_2>k_1} 2a_t \frac{1}{\lambda_t-\lambda_{k_1}}\frac{1}{\lambda_t-\lambda_{k_2}}. \tag{4.6}$$

Note that

$$\begin{aligned} b_t^2 &= \sum_{k_1\neq t}\frac{1}{\lambda_{k_1}-\lambda_t}\sum_{k_2\neq t}\frac{1}{\lambda_{k_2}-\lambda_t} \\ &= \sum_{k_1\neq t}\frac{1}{(\lambda_{k_1}-\lambda_t)^2} - \sum_{k_1\neq t}\sum_{k_2\neq k_1}\frac{1}{\lambda_{k_1}-\lambda_t}\frac{1}{\lambda_{k_2}-\lambda_t} \\ &= g_t - 2\sum_{k_1\neq t}\sum_{k_2>k_1}\frac{1}{\lambda_{k_1}-\lambda_t}\frac{1}{\lambda_{k_2}-\lambda_t}. \end{aligned}$$

So,

$$2\sum_{k_1\neq t}\sum_{k_2>k_1}\frac{1}{\lambda_t-\lambda_{k_1}}\frac{1}{\lambda_t-\lambda_{k_2}} = g_t - b_t^2. \tag{4.7}$$

Hence, (4.6) can be written as $(g_t - b_t^2)a_t$.
Thus, $S(m_t)$ is given by

$$S(m_t) = S(2) = -a_t - 2a_t b_t - a_t b_t^2 - a_t g_t. \tag{4.8}$$

Therefore,

$$\begin{aligned} \frac{\partial^2}{\partial\lambda_t^2}\big(c(\Lambda)\big) &= (-1)^n 2\pi^p \sum_{i=1}^{p}(-1)_i^m \exp(-\lambda_i) S(m_i) \\ &= 2\pi^p\Big\{-\sum_{i\neq t}\frac{2a_i e^{\lambda_i}}{(\lambda_t-\lambda_i)^2} + a_t e^{\lambda_t}\big[1+g_t-2b_t+b_t^2\big]\Big\}. \end{aligned} \tag{4.9}$$

Since equation (4.9) can be found in Amaral, Florez and Cysneiros (2013), it illustrates the equivalence between the results of Kume and Wood (2007) and Amaral, Florez and Cysneiros (2013).

5 Saddlepoint approximation for the Bingham normalizing constants

One can use a saddlepoint approximation to the Bingham normalizing constant (Kume and Wood, 2005). The normalizing constant can be written as

$$\hat{c}_3(\lambda) = 2^{1/2}\pi^{(p-1)}\{K_\lambda^{(2)}(\hat{t})\}^{-1/2}\exp\{T(\hat{t})-\hat{t}\}\prod_{i=1}^{p}(\lambda_i-\hat{t})^{-1/2}, \tag{5.1}$$

where $K_\lambda(t) = -\frac{1}{2}\sum_{i=1}^{p}\log(1-t/\lambda_i)$, $K_\lambda^{(j)} = \frac{\partial^j K_\lambda(t)}{\partial t^j} = \sum_{i=1}^{p}\frac{(j-1)!}{2(\lambda_i-t)^j}$, $j \geq 1$, the term $\hat{t}$ is the unique solution in $(-\infty, \min_i \lambda_i)$ to the equation $K_\lambda^{(1)} = 1$ and $T(t) = \frac{1}{8}\rho_4(t) - \frac{5}{24}\rho_3(t)^2$, where $\rho_j(t) = \frac{K_\lambda^{(j)}}{\{K_\lambda^{(2)}(t)\}^{j/2}}$.

6 Bootstrap bias correction

The bootstrap method was initially proposed by Efron (1979). The method's idea is simple. Since the sampling distribution of a statistics is often unknown, samples with replacement from the original sample can be used to estimate that.

The bootstrap method can be parametric or nonparametric. The parametric bootstrap assumes a distribution for the data. On the other hand, the nonparametric bootstrap does not assume any distribution.

The nonparametric bootstrap method for bias correction has been extensively used in the literature (Efron, 1979). This method is briefly explained in this section.

Let $z_1, \dots, z_n$ be a random sample from the density (2.3). The bootstrap bias correction method has the following steps:

1. Obtain the maximum likelihood estimate (MLE) $\hat{\lambda}$ of λ by maximizing (2.4).
2. Choose B random samples with replacement from $z_1, \dots, z_n$.
3. Compute the MLE of each bootstrap sample, which will deliver
$$\hat{\lambda}^1, \dots, \hat{\lambda}^B.$$
4. Calculate the mean of the bootstrap samples $\bar{\lambda} = B^{-1}\sum_{i=1}^{B}\lambda^i$.
5. The bootstrap bias corrected estimator is given by
$$\tilde{\lambda} = 2\hat{\lambda} - \bar{\lambda}. \tag{6.1}$$

The parametric bootstrap method for bias correction is similar to the previous one. However, since a probabilistic model is assumed, the samples with replacement are draw from the model (2.3). The parameters are replaced by the MLE $\hat{\lambda}$ of λ.

The parametric bootstrap method can be described as follows:

- Consider $z_1, \dots, z_n$ a random sample from (2.3);
- Using the model (2.3), estimate by maximum likelihood the parameter vector λ and the parameter matrix A;
- Draw B samples with replacement from $f(z) = c(\hat{A})^{-1}\exp(z^\star \hat{A} z)$, $z \in \mathbb{C}S^{k-1}$, where $\hat{A}$ is the MLE of A. Let each bootstrap sample be denoted by $z_1^{(b)}, \dots, z_n^{(b)}$;
- For each bootstrap sample, compute $\hat{\lambda}^b$, which is the MLE of λ for the sample $z_1^{(b)}, \dots, z_n^{(b)}$;

- The bias corrected estimator is computed such as the previous method. So

$$\tilde{\lambda} = 2\hat{\lambda} - \bar{\lambda}, \tag{6.2}$$

where $\bar{\lambda} = \sum_{i=1}^{b} \hat{\lambda}^{b}$ and $\hat{\lambda}$ is the MLE of λ.

7 Numerical evaluation

The methods were defined according to the list below:

- MLE: maximum likelihood estimation
  It is obtained by maximizing (2.4).
- MLE-SA: maximum likelihood estimation with saddlepoint approximation
  The normalizing constant (5.1) is used.
- BC-MLE: bias correction of the maximum likelihood estimates
  The formula (3.1) is applied.
- BC-MLE-SA: bias correction of the maximum likelihood estimates with sadllepoint approximation
  The formulas (5.1) and (3.1) are implemented.
- Boot-NPAR: bias correction with nonparametric bootstrap
  The estimator (6.1) is calculated.
- Boot-PAR: bias correction with parametric bootstrap
  The statistics (6.2) is used.
- Boot-SA-NPAR: bias correction with saddlepoint approximation and nonparametric bootstrap
  The normalizing constant (5.1) and the estimator (6.1) are computed.
- Boot-SA-PAR: bias correction with saddlepoint approximation and parametric bootstrap
  The normalizing constant (5.1) and the estimator (6.2) are calculated.

The results of simulation experiments with 10,000 Monte Carlo samples and 1000 random bootstrap samples are shown in the next tables. The parameter vectors of the complex Bingham distribution are defined according to following experiment: $\lambda = \sigma \times (4, 3, 2, 1, 0)$, where $\sigma = 10$ and the sample sizes were 20 and 60. Thus, the concentration is proportional to the value of σ. The experiments were implemented in R.

The results of Table 1 can be summarized as follows. The analytical bias correction improves the MSE in all considered cases of λ. On the other hand, the bootstrap methods reduces the MSE only for the two bigger values of λ. The methods reduce the bias only for the first three values of λ, and the bias correction methods do not reduce the bias when $\lambda_4 = 10$. The results also indicate that the saddlepoint approximation did not change the previous conclusions. It is relevant compare the nonparametric bootstrap method to the parametric bootstrap method. The parametric bootstrap method has smaller bias and MSE when $\lambda_1 = 40$. However, for other values of λ, they have a similar performance.

Table 1 *Simulation results for σ equals to* 10 *and the sample size equals to* 20

Estimator	$\lambda_1 = 40$			$\lambda_2 = 30$			$\lambda_3 = 20$			$\lambda_4 = 10$		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
MLE	−26.32	263.08	955.91	−6.45	47.55	89.17	−0.98	15.00	15.96	0.02	4.38	4.38
MLE-SA	−26.31	263.17	955.21	−6.43	47.61	88.94	−0.94	15.07	15.96	0.10	4.44	4.45
BC-MLE	−23.01	237.42	766.80	−4.63	42.91	64.37	0.06	13.52	13.53	0.50	3.91	4.17
BC-MLE-SA	−22.93	237.09	762.87	−4.57	42.90	63.82	0.12	13.58	13.59	0.59	3.97	4.32
Boot-NPAR	11.09	156.25	279.16	2.52	57.71	64.05	0.79	21.29	21.92	0.07	5.77	5.77
Boot-PAR	7.08	149.36	199.45	2.76	55.50	63.14	0.92	21.18	22.03	0.07	5.81	5.82
Boot-SA-NPAR	11.11	156.34	279.66	2.54	57.79	64.26	0.83	21.39	22.09	0.15	5.85	5.88
Boot-SA-PAR	7.10	149.45	199.80	2.79	55.58	63.37	0.96	21.28	22.21	0.15	5.90	5.92

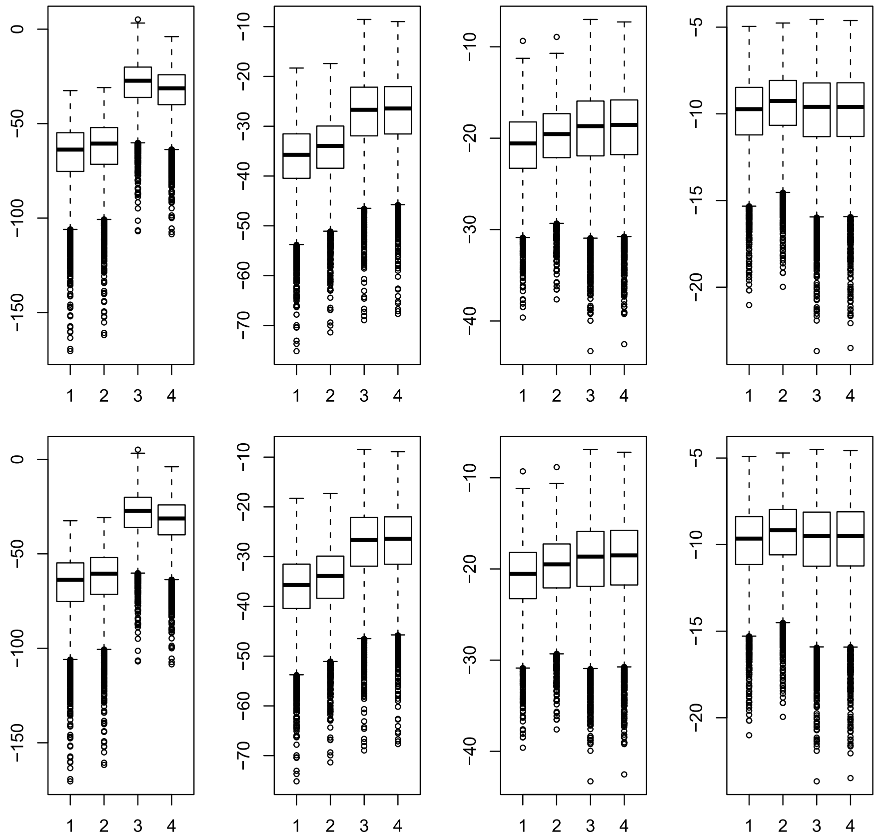

Figure 1 *Boxplots for eight bias correction methods* ($\sigma = 10$ *and sample size of* 20).

The boxplots of Figure 1 are useful to evaluate the dispersion of the results. The methods with smallest dispersion are the BC-MLE-SA and Boot-SA-PAR. It indicates that the saddlepoint approximation is key to reduce dispersion.

In Table 2, the simulation results for the sample size of 60 and $\sigma = 10$ are shown. Since the sample size is larger, all the biases and MSE are smaller that those of Table 1. It is useful to suggest that the analytical and bootstrap methods should converge for similar biases and MSE. Since the dominant eigenvalue is used to obtain the mean shape, which is a relevant statistics in shape analysis, we focus on this quantity. So the results indicate that the bootstrap methods have smaller bias and MSE. The parametric and nonparametric bootstrap methods deliver almost the same results.

The boxplots of all methods are shown in Figure 2. As in the previous graph, the methods with smallest dispersion are the BC-MLE-SA and Boot-SA-PAR. Again, the dispersion reduction of the saddlepoint methods is noticeable.

Table 2 *Simulation results for σ equals to* 10 *and the sample size equals to* 60

	$\lambda_1 = 40$			$\lambda_2 = 30$			$\lambda_3 = 20$			$\lambda_4 = 10$		
Estimator	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
MLE	−6.99	35.90	84.73	−1.79	13.43	16.62	−0.27	5.76	5.83	−0.03	1.66	1.66
MLE-SA	−6.97	35.93	84.50	−1.76	13.45	16.55	−0.23	5.79	5.84	0.05	1.69	1.69
BC-MLE	−6.21	34.71	73.23	−1.26	12.98	14.56	0.07	5.57	5.57	0.13	1.60	1.62
BC-MLE-SA	−6.17	34.72	72.82	−1.22	13.00	14.49	0.11	5.60	5.61	0.21	1.63	1.67
Boot-NPAR	0.48	35.46	35.69	0.30	17.75	17.84	0.07	7.51	7.51	−0.03	1.88	1.88
Boot-PAR	0.45	35.38	35.58	0.34	17.76	17.88	0.08	7.53	7.53	−0.03	1.88	1.88
Boot-SA-NPAR	0.50	35.49	35.74	0.32	17.79	17.89	0.11	7.55	7.56	0.05	1.91	1.91
Boot-SA-PAR	0.47	35.41	35.63	0.36	17.79	17.93	0.12	7.56	7.58	0.05	1.91	1.91

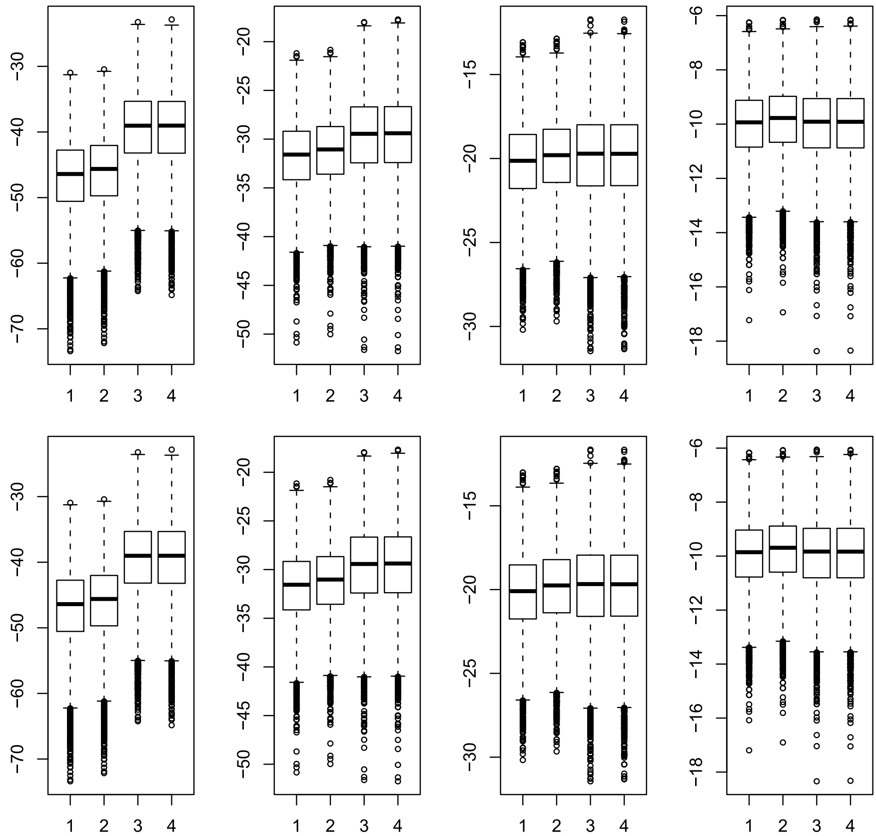

Figure 2 *Boxplots for eight bias correction methods* ($\sigma = 10$ *and sample size of* 60).

8 Conclusions

Some bias correction methods for the complex Bingham distribution has been derived. The analytical and bootstrap methods have reduced the mean square error of the maximum likelihood estimator. For the dominant eigenvalue, the bootstrap bias corrected estimator has slightly smaller mean square error than the analytical bias corrected estimator. When the sample size increases, the biases of all methods goes to zero. Moreover, the saddlepoint approximation is useful to reduce the dispersion of the results.

Appendix: Terms of the second-order bias correction

After some algebraic operations, we have the following results:

$$c_{rr}(\Lambda) = (2\pi)^{k-1}\left[2\sum_{j\neq r}\frac{u_{jr}}{\lambda_j - \lambda_r} + a_r e^{\lambda_r}\left(1 - 2b_r + b_r^2 + g_r\right)\right]$$

and

$$c_{rs}(\Lambda) = (2\pi)^{k-1}\Bigg[2\sum_{\substack{j\neq r\\ j\neq s}} \frac{u_{js}}{\lambda_j - \lambda_r} + u_{rs}\Big(1 - \frac{1}{\lambda_r - \lambda_s} - b_r\Big)$$

$$+ u_{sr}\Big(1 - \frac{1}{\lambda_s - \lambda_r} - b_s\Big)\Bigg],$$

where $u_{rs} = \frac{a_r e^{\lambda_s}}{\lambda_r - \lambda_s}$, $b_r = \sum_{j\neq r} \frac{1}{(\lambda_r - \lambda_j)}$ and $g_r = \sum_{j\neq r} \frac{1}{(\lambda_r - \lambda_j)^2}$. See Amaral, Florez and Cysneiros (2013).

We also have

$$c_{rrr}(\Lambda) = 2\pi^{k-1}\Bigg\{6\sum_{j\neq r} \frac{u_{jr}}{(\lambda_j - \lambda_r)^2}$$

$$+ a_r e^{\lambda_r}\Big[(1 - b_r)(1 - 2b_r + b_r^2 + g_r) + 2g_r\Big(1 - b_r - \frac{h_r}{g_r}\Big)\Big]\Bigg\},$$

where $h_r = \sum_{j\neq r} \frac{1}{(\lambda_r - \lambda_j)^3}$;

$$c_{rrt}(\Lambda) = 2\pi^{k-1}\Bigg\{2\sum_{\substack{j\neq r\\ j\neq t}} \frac{u_{jr}}{(\lambda_j - \lambda_r)(\lambda_j - \lambda_t)} + \frac{2u_{tr}}{\lambda_t - \lambda_r}\Big(1 - b_t - \frac{2}{\lambda_t - \lambda_r}\Big)$$

$$+ u_{rt}\Big[1 - 2b_r + b_r^2 + g_r + \frac{2}{\lambda_r - \lambda_t}\Big(\frac{1}{\lambda_r - \lambda_t} + b_r - 1\Big)\Big]\Bigg\},$$

$$c_{rsr}(\Lambda) = 2\pi^{k-1}\Bigg\{2\sum_{\substack{j\neq r\\ j\neq s}} \frac{u_{jr}}{(\lambda_j - \lambda_r)(\lambda_j - \lambda_s)}$$

$$+ u_{rs}\Big[\Big(1 - \frac{1}{\lambda_r - \lambda_s} - b_r\Big)^2 + \frac{1}{(\lambda_r - \lambda_s)^2} + g_r\Big]$$

$$+ \frac{2u_{sr}}{\lambda_s - \lambda_r}\Big(1 - \frac{2}{\lambda_s - \lambda_r} - b_s\Big)\Bigg\},$$

$$c_{rss}(\Lambda) = 2\pi^{k-1}\Bigg\{2\sum_{\substack{j\neq r\\ j\neq s}} \frac{u_{js}}{(\lambda_j - \lambda_r)(\lambda_j - \lambda_s)}$$

$$+ u_{sr}\Big[\Big(1 - \frac{1}{\lambda_s - \lambda_r} - b_s\Big)^2 + \frac{1}{(\lambda_s - \lambda_r)^2} + g_s\Big]$$

$$+ \frac{2u_{rs}}{\lambda_r - \lambda_s}\Big(1 - \frac{2}{\lambda_r - \lambda_s} - b_r\Big)\Bigg\}$$

and

$$
\begin{aligned}
c_{rst}(\Lambda) = 2\pi^{k-1}\Bigg\{ & \sum_{\substack{j\neq r\\ j\neq s\\ j\neq t}} \frac{u_{jr}}{(\lambda_j-\lambda_s)(\lambda_j-\lambda_t)} \\
& + \frac{u_{tr}}{\lambda_t-\lambda_s}\left[1-\frac{1}{\lambda_t-\lambda_r}-\frac{1}{\lambda_t-\lambda_s}-b_t\right] \\
& + \frac{u_{rs}}{\lambda_r-\lambda_t}\left(1-\frac{1}{\lambda_r-\lambda_s}-\frac{1}{\lambda_r-\lambda_t}-b_r\right) \\
& + \frac{u_{sr}}{\lambda_s-\lambda_t}\left(1-\frac{1}{\lambda_s-\lambda_r}-\frac{1}{\lambda_s-\lambda_t}-b_s\right)\Bigg\}.
\end{aligned}
$$

Acknowledgments

The research of Getulio J. A. Amaral is supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil (CNPq-Brazil) and FACEPE. We also thank the anonymous referees for their valuable comments.

References

Amaral, G. J. A., Florez, O. P. R. and Cysneiros, F. J. A. (2013). Graphical and numerical methods for detecting influential observations in complex Bingham data. *Communications in Statistics. Simulation and Computation* **42**, 1801–1814. MR3042800

Bartlett, M. S. (1953). Approximate confidence intervals. *Biometrika* **40**, 12–19. MR0056889

Bingham, C. (1974). An antipodally symmetric distribution on the sphere. *The Annals of Statistics* **2**, 1201–1225. MR0397988

Cordeiro, G. M. and Cribari-Neto, F. (2014). *An Introduction to Bartlett Correction and Bias Reduction*. Heidelberg: Springer. MR3289997

Cordeiro, G. M. and Klein, R. (1994). Bias correction in ARMA models. *Statistics & Probability Letters* **19**, 169–176. MR1278646

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **30**, 248–275. MR0237052

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Chichester: John Wiley and Sons. MR1646114

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics* **7**, 1–26. MR0515681

Kent, J. T. (1994). The complex Bingham distribution and shape analysis. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **56**, 285–299. MR1281934

Kent, J. T., Constable, P. D. L. and Er, F. (2004). Simulation of the complex Bingham distribution. *Statistics and Computing* **14**, 53–57. MR2019905

Kume, A. and Wood, A. T. A. (2005). Saddlepoint approximations for the Bingham and Fisher–Bingham normalising constants. *Biometrika* **92**, 465–476. MR2201371

Kume, A. and Wood, A. T. A. (2007). On the derivatives of the normalising constant of the Bingham distribution. *Statistics & Probability Letters* **77**, 832–837. MR2369690

Mardia, K. I. and Jupp, P. E. (2000). *Directional Statistics*. Chichester: Wiley. MR1828667

Small, C. G. (1996). *The Statistical Theory of Shape*. New York: Springer. MR1418639

Watson, G. S. (1965). Equatorial distributions on a sphere. *Biometrika* **52**, 193–201. MR0207115

L. H. G. Dore
Departamento de Estatística e Ciências Atuariais
Universidade Federal de Sergipe
Cidade Universitária Professor José Aloisio Campos
Av. Marechal Rondon s/n
Jardin Rosa Else
São Cristovão, SE 49100-000
Brasil
E-mail: luizgd@yahoo.com.br

G. J. A. Amaral
Departamento de Estatística
Universidade Federal de Pernambuco
Cidade Universitária
Recife, PE 50740-540
Brasil
E-mail: gjaa@de.ufpe.br

J. T. M. Cruz
Departamento de Ciências Contábeis e Atuarias
Universidade Federal de Pernambuco
Av. Prof. Moraes Rego 1235
Cidade Universitária
Recife, PE 50670-901
Brasil
E-mail: tiagocarpdien@gmail.com

A. T. A. Wood
School of Mathematical Sciences
University of Nottingham
Room C10
The Mathematical Sciences Building
University Park, Nottingham NG7 2RD
United Kingdom
E-mail: atw@maths.nottingham.ac.uk