Journal of Multivariate Analysis 101 (2010) 2091–2102

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

The multivariate Behrens–Fisher distribution

Fco. Javier Girón *, Carmen del Castillo

Department of Statistics and Operations Research, Faculty of Sciences, University of Malaga, Campus de Teatinos, 29071 Malaga, Spain

ARTICLE INFO

Article history:
Received 27 June 2009
Available online 24 April 2010

AMS subject classifications:
60E05
62C10
62F15
65C05

Keywords:
Multivariate Behrens–Fisher distribution
Convolution
Mixture
H.P.D. regions
Monte Carlo methods
Frequentist coverage

ABSTRACT

The main purpose of this paper is the study of the multivariate Behrens–Fisher distribution. It is defined as the convolution of two independent multivariate Student t distributions. Some representations of this distribution as the mixture of known distributions are shown. An important result presented in the paper is the elliptical condition of this distribution in the special case of proportional scale matrices of the Student t distributions in the defining convolution. For the bivariate Behrens–Fisher problem, the authors propose a non-informative prior distribution leading to highest posterior density (H.P.D.) regions for the difference of the mean vectors whose coverage probability matches the frequentist coverage probability more accurately than that obtained using the independence-Jeffreys prior distribution, even with small samples.

1. Introduction

The multivariate Behrens–Fisher problem [1,10,11] consists of testing whether the means of two independent multivariate normal distributions are the same, when the covariance matrices are unequal and unknown.

There is a vast literature devoted to the solution of this problem. Some approximate and exact classical solutions were suggested by Bennet [2], James [17], Yao [29], Johansen [18], Nel et al. [24,25], Kim [20], Krishnamoorthy and Yu [21], Gamage et al. [12], Yanagihara and Yuan [28] and Buot et al. [4,5].

Some Bayesian solutions, in the univariate case, were illustrated by Box and Tiao [3], Girón et al. [16] and Moreno et al. [22].

Johnson and Weerahandi [19], Nel and Groenewald [23] and Thabane and Safiul Haq [27] are among those who used the Bayesian approach to study the problem.

On the other hand, the multivariate Behrens–Fisher distribution was defined by Dickey [8] as the convolution of a finite number of multivariate Student t distributions. Depending on the structure of the scale matrices of the terms of the convolution, we can define two types of multivariate Behrens–Fisher distributions. For one of the types, we investigate the elliptical condition of these distributions, because in this case we could obtain an exact Bayesian solution of the multivariate Behrens–Fisher problem based on highest posterior density (H.P.D.) regions – which are ellipsoidal – easily. However, we will show that the credible regions of the multivariate Behrens–Fisher distribution are not ellipsoidal in general, and in some cases they are not even convex sets. This implies that an exact analytical solution to this problem is difficult to obtain, and, for this reason, we have to resort to numerical or Monte Carlo methods.

* Corresponding author.
E-mail address: fj_giron@uma.es (F.J. Girón).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.04.008

The assumption of proportional scale matrices of the Student t distributions of the convolution is a necessary one to state the elliptical condition of the Behrens–Fisher distribution. This is the reason we have considered, in Section 2, two types of multivariate Behrens–Fisher distribution, type 1, for arbitrary scale matrices, and type 2, for the case of proportional scale matrices.

In Section 3, some representations of both types of the multivariate Behrens–Fisher distribution are given. We represent them as location mixture of a Student t distribution when the mixing distribution is, in turn, a Student t, proved in Girón et al. [16] for the univariate case, and as a scale mixture of Student t when the mixing distribution is an inverted beta 2. We also provide a particular case of the multivariate Behrens–Fisher distribution of type 2 as a finite mixture of Student t distributions, under some particular conditions.

The elliptical condition of the multivariate Behrens–Fisher distribution of type 2 is studied in Section 4.

In Section 5, the frequentist behavior of the H.P.D. regions for the multivariate Behrens–Fisher problem where the parameter of interest is the difference of the mean vectors are examined. A non-informative prior distribution over the mean vectors and the covariance matrices of the two multivariate normal populations of the problem is considered that leads to H.P.D. regions for the difference of the mean vectors whose coverage probability matches the frequentist coverage probability more accurately than that obtained using the independence-Jeffreys prior distribution. For the case of the bivariate Behrens–Fisher problem, the simulation results indicate excellent matching even with small samples.

Finally, a discussion of the results of the work is presented is Section 6.

2. Definitions

In this section, the multivariate Behrens–Fisher distribution is defined as the convolution of two multivariate Student t distributions, which is a generalization of the definition that appears in Box and Tiao [3] for the univariate case.

As was explained in Section 1, two types of multivariate Behrens–Fisher distribution are considered: type 1, the general case, in which there are no restrictions on the scale matrices of the Student t distributions in the convolution; and type 2, the particular case of proportional scale matrices.

Definition 2.1. A random vector $\mathbf{x} = (x_1, \ldots, x_k)'$ follows a multivariate standard Behrens–Fisher distribution of type 1 with parameters $\mathbf{A}_1, \mathbf{A}_2, \nu_1$ and ν_2 ($\mathbf{A}_1$ and $\mathbf{A}_2$ non-singular $k \times k$ matrices satisfying $\mathbf{A}_1\mathbf{A}_1' + \mathbf{A}_2\mathbf{A}_2' = \mathbf{I}_k$ and $\nu_1, \nu_2 > 0$) if $\mathbf{x} \overset{\mathrm{d}}{=} \mathbf{A}_1\mathbf{t}_1 - \mathbf{A}_2\mathbf{t}_2$, where $\mathbf{t}_1$ and $\mathbf{t}_2$ are independent random vectors following multivariate t distributions, $\mathbf{t}_i \sim \mathrm{t}_k(\mathbf{t}_i|\mathbf{0}, \mathbf{I}_k, \nu_i)$ for $i = 1, 2$.

It will be denoted as $\mathbf{x} \sim \text{Be–Fi}_k^1(\mathbf{x}|\mathbf{A}_1, \mathbf{A}_2, \nu_1, \nu_2)$.

Definition 2.2. A random vector $\mathbf{x} = (x_1, \ldots, x_k)'$ follows a multivariate standard Behrens–Fisher distribution of type 2 with parameters ϕ, ν_1 and ν_2 ($\phi \in \left[0, \frac{\pi}{2}\right]$ and $\nu_1, \nu_2 > 0$) if $\mathbf{x} \overset{\mathrm{d}}{=} \cos\phi\, \mathbf{t}_1 - \sin\phi\, \mathbf{t}_2$, where $\mathbf{t}_1$ and $\mathbf{t}_2$ are independent random vectors and $\mathbf{t}_i \sim \mathrm{t}_k(\mathbf{t}_i|\mathbf{0}, \mathbf{I}_k, \nu_i)$ for $i = 1, 2$.

It will be denoted as $\mathbf{x} \sim \text{Be–Fi}_k^2(\mathbf{x}|\phi, \nu_1, \nu_2)$.

The extension of these distributions to a location-scale family is the usual one.

Definition 2.3. A random vector $\mathbf{y} = (y_1, \ldots, y_n)'$ follows a multivariate Behrens–Fisher distribution of type 1 with parameters $\boldsymbol{\mu}, \mathbf{B}, \mathbf{A}_1, \mathbf{A}_2, \nu_1$ and ν_2 ($\boldsymbol{\mu}$ a $n \times 1$ vector, $\mathbf{B}$ a $n \times k$ matrix, $n \le k$, $\mathbf{A}_1$ and $\mathbf{A}_2$ non-singular $k \times k$ matrices satisfying $\mathbf{A}_1\mathbf{A}_1' + \mathbf{A}_2\mathbf{A}_2' = \mathbf{I}_k$ and $\nu_1, \nu_2 > 0$) if $\mathbf{y} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{B}\mathbf{x}$ where $\mathbf{x} \sim \text{Be–Fi}_k^1(\mathbf{x}|\mathbf{A}_1, \mathbf{A}_2, \nu_1, \nu_2)$.

It will be denoted as $\mathbf{y} \sim \text{Be–Fi}_n^1(\mathbf{y}|\boldsymbol{\mu}, \mathbf{B}\mathbf{A}_1, \mathbf{B}\mathbf{A}_2, \nu_1, \nu_2)$.

Definition 2.4. A random vector $\mathbf{y} = (y_1, \ldots, y_n)'$ follows a multivariate Behrens–Fisher distribution of type 2 with parameters $\boldsymbol{\mu}, \mathbf{B}, \phi, \nu_1$ and ν_2 ($\boldsymbol{\mu}$ a $n \times 1$ vector, $\mathbf{B}$ a $n \times k$ matrix, $n \le k$, $\phi \in \left[0, \frac{\pi}{2}\right]$ and $\nu_1, \nu_2 > 0$) if $\mathbf{y} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{B}\mathbf{x}$ where $\mathbf{x} \sim \text{Be–Fi}_k^2(\mathbf{x}|\phi, \nu_1, \nu_2)$.

It will be denoted as $\mathbf{y} \sim \text{Be–Fi}_n^2(\mathbf{y}|\boldsymbol{\mu}, \mathbf{B}, \phi, \nu_1, \nu_2)$.

By examining the above definitions, it is easily deduced that the multivariate Behrens–Fisher distribution of type 2 is a subclass of the type1 class of distributions. It is also obtained that

If $\mathbf{x} \overset{\mathrm{d}}{=} t_n(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1) - t_n(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2)$, then

$$\mathbf{x} \sim \text{Be–Fi}_n^1(\mathbf{x}|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1^{\frac{1}{2}}, \boldsymbol{\Sigma}_2^{\frac{1}{2}}, \nu_1, \nu_2), \tag{1}$$

where $\boldsymbol{\Sigma}_i^{\frac{1}{2}}$ is a positive definite matrix such that $\boldsymbol{\Sigma}_i^{\frac{1}{2}}\boldsymbol{\Sigma}_i^{\frac{1}{2}} = \boldsymbol{\Sigma}_i$.

If $\mathbf{x} \overset{\mathrm{d}}{=} t_n(\boldsymbol{\mu}_1, a\boldsymbol{\Sigma}, \nu_1) - t_n(\boldsymbol{\mu}_2, b\boldsymbol{\Sigma}, \nu_2)$, then

$$\mathbf{x} \sim \text{Be–Fi}_n^2\left(\mathbf{x}|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \sqrt{a+b}\,\boldsymbol{\Sigma}^{\frac{1}{2}}, \phi, \nu_1, \nu_2\right), \tag{2}$$

where ϕ is an angle in $\left[0, \frac{\pi}{2}\right]$ such that $\phi = \tan^{-1}\left(\sqrt{\frac{b}{a}}\right)$.

3. Some characterizations and needed results

A random variable X follows an inverted-gamma distribution with parameters α and β, and will be denoted as $X \sim \mathrm{Ga}^{-1}(\alpha, \beta)$, if its reciprocal $1/X$ follows a gamma distribution, $\mathrm{Ga}(\alpha, \beta)$, with shape parameter α and scale parameter β.

Its density function is

$$f(x|\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-(\alpha+1)} \mathrm{e}^{-\beta/x} \quad \text{if } x > 0.$$

A random matrix $\mathbf{X}$ follows an inverted Wishart distribution with parameters $\mathbf{A}$ and ν ($\mathbf{A}$ a positive definite $k \times k$ matrix and $\nu > k - 1$) if its inverse $\mathbf{X}^{-1}$ follows a Wishart distribution, $\mathrm{W}_k(\mathbf{A}, \nu)$, with matrix $\mathbf{A}$ and ν degrees of freedom.

Its density function is

$$f(\mathbf{X}|\mathbf{A}, \nu) = \left(\frac{1}{2}\right)^{\frac{k\nu}{2}} \left(\Gamma\left(\frac{1}{2}\right)\right)^{\frac{k(k-1)}{2}} \left(\prod_{i=1}^{k} \Gamma\left(\frac{\nu+i-k}{2}\right)\right)^{-1} |\mathbf{A}|^{\frac{\nu}{2}} |\mathbf{X}|^{-\frac{\nu+k+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{X}^{-1}\mathbf{A})\right),$$

where $\mathbf{X}$ is a symmetric and positive definite $k \times k$ matrix. These inverted distributions appear in Bayesian inference in a natural way, as the posterior distribution of the variance and covariance matrix in univariate and multivariate normal sampling, respectively, when reference or conjugate distributions on the parameters are used.

They also appear in the following well known representation of the multivariate Student t distribution as a scale mixture of multivariate normals when the mixing distribution is inverted gamma or inverted Wishart.

In fact, if $\mathbf{t} \sim \mathrm{t}_k(\mathbf{t}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, then

$$\mathbf{t} \sim \int \mathrm{N}_k(\mathbf{t}|\boldsymbol{\mu}, \lambda\boldsymbol{\Sigma}) \mathrm{dGa}^{-1}\left(\lambda \left| \frac{\nu}{2}, \frac{\nu}{2}\right.\right), \tag{3}$$

$$\mathbf{t} \sim \int \mathrm{N}_k(\mathbf{t}|\boldsymbol{\mu}, \boldsymbol{\Lambda}) \mathrm{dW}_k^{-1}(\boldsymbol{\Lambda}|\nu\boldsymbol{\Sigma}, \nu + k - 1).$$

The following theorem provides a representation of the multivariate Behrens–Fisher distribution which is a consequence of this mixture distribution.

Theorem 3.1. *If* $\mathbf{x} = (x_1, \ldots, x_k)' \sim \text{Be–Fi}_k^1(\mathbf{x}|\mathbf{A}_1, \mathbf{A}_2, \nu_1, \nu_2)$, *then*

$$\mathbf{x} \sim \iint \mathrm{N}_k(\mathbf{x}|\mathbf{0}, \lambda_1\mathbf{A}_1\mathbf{A}_1' + \lambda_2\mathbf{A}_2\mathbf{A}_2') \prod_{i=1}^{2} \mathrm{dGa}^{-1}\left(\lambda_i \left| \frac{\nu_i}{2}, \frac{\nu_i}{2}\right.\right)$$

$$\mathbf{x} \sim \iint \mathrm{N}_k(\mathbf{x}|\mathbf{0}, \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2) \prod_{i=1}^{2} \mathrm{dW}_k^{-1}(\boldsymbol{\Lambda}_i|\nu_i\mathbf{A}_i\mathbf{A}_i', \nu_i + k - 1).$$

As a consequence of Theorem 3.1, the following representation for the multivariate Behrens–Fisher distribution of type 2 is obtained.

If $\mathbf{x} = (x_1, \ldots, x_k)' \sim \text{Be–Fi}_k^2(\mathbf{x}|\phi, \nu_1, \nu_2)$ then, it can be expressed as

$$\begin{aligned}
\mathbf{x} &\sim \iint \mathrm{N}_k\left(\mathbf{x}|\mathbf{0}, (\lambda_1\cos^2\phi + \lambda_2\sin^2\phi)\mathbf{I}_k\right) \prod_{i=1}^{2} \mathrm{dGa}^{-1}\left(\lambda_i \left| \frac{\nu_i}{2}, \frac{\nu_i}{2}\right.\right) \\
&\sim \iint \mathrm{N}_k\left(\mathbf{x}|\mathbf{0}, (x_1 + x_2)\mathbf{I}_k\right) \mathrm{dGa}^{-1}\left(x_1 \left| \frac{\nu_1}{2}, \frac{\nu_1}{2}\cos^2\phi\right.\right) \mathrm{dGa}^{-1}\left(x_2 \left| \frac{\nu_2}{2}, \frac{\nu_2}{2}\sin^2\phi\right.\right) \\
&\sim \int \mathrm{N}_k\left(\mathbf{x}|\mathbf{0}, z\mathbf{I}_k\right) \mathrm{dCGa}^{-1}\left(z \left| \frac{\nu_1}{2}, \frac{\nu_1}{2}\cos^2\phi, \frac{\nu_2}{2}, \frac{\nu_2}{2}\sin^2\phi\right.\right),
\end{aligned}$$

where $\mathrm{CGa}^{-1}(\alpha_1, \beta_1, \alpha_2, \beta_2)$ denotes the convolution of two inverted-gamma distributions with parameters (α_1, β_1) and (α_2, β_2), respectively.

The last formula provides the connection between the Behrens–Fisher distribution and the convolution of two inverted gamma distributions. It essentially shows that the Behrens–Fisher distribution is a scale mixture of normals when the mixing distribution is a convolution of two inverted gammas.

Although in general, the convolution of two inverted gamma densities does not have an explicit or simple form, the next theorem, proved in Girón and Castillo [15], shows that, under some restrictions on the shape parameters, the convolution of inverted gamma distributions is distributed as a finite mixture of inverted gamma distributions all having the same scale parameter.

Theorem 3.2. *If $x \sim \mathrm{Ga}^{-1}\left(n+\frac{1}{2}, \beta_1\right)$, $y \sim \mathrm{Ga}^{-1}\left(m+\frac{1}{2}, \beta_2\right)$, $n, m \in \mathbb{N}$, $m \geq n$ and x and y are independent, then the convolution of x and y is distributed as the following mixture*

$$x+y \sim \sum_{i=1}^{m+1} p_i \,\mathrm{Ga}^{-1}\left(n-\frac{1}{2}+i, \left(\sqrt{\beta}_1+\sqrt{\beta}_2\right)^2\right),$$

where the weights $p_i \geq 0$, $i = 1, \dots, m+1$, $\sum_{i=1}^{m+1} p_1 = 1$ are computed in a recursive manner from the formulae

$$p_{m+1} = \frac{\sqrt{\pi}\,\Gamma\left(n+m+\frac{1}{2}\right)}{\Gamma\left(n+\frac{1}{2}\right)\Gamma\left(m+\frac{1}{2}\right)} \frac{\left(\sqrt{\beta_1}\right)^n \left(\sqrt{\beta_2}\right)^m}{\left(\sqrt{\beta_1}+\sqrt{\beta_2}\right)^{n+m}}$$

$$p_{j+1} = 2^{2j+2}\Gamma\left(n+\frac{1}{2}+j\right)\left(\frac{c\gamma_{n+j}}{2^{2+2j}\sqrt{\pi}\left(\sqrt{\beta_1}+\sqrt{\beta_2}\right)^{n+j}} - \sum_{i=j+2}^{m+1} \frac{p_i}{2^{2i}\Gamma\left(n-\frac{1}{2}+i\right)} \cdot \frac{(n+2i-2-j)!}{(n+j)!(i-1-j)!}\right),$$
$$j = 0, \dots, m-1,$$

where

$$c = \frac{\pi}{2^{2(n+m)}\Gamma\left(n+\frac{1}{2}\right)\Gamma\left(m+\frac{1}{2}\right)};$$

$$\gamma_k = 2^{2k}\sum_{i=0}^{k} \frac{(2n-i)!}{i!(n-i)!} \frac{(2m-k+i)!}{(k-i)!(m-k+i)!} \left(\sqrt{\beta_1}\right)^i \left(\sqrt{\beta_2}\right)^{k-i} \quad \text{for } k = 0, \dots, n;$$

$$\gamma_{n+k} = 2^{2(n+k)}\sum_{i=0}^{n} \frac{(2n-i)!}{i!(n-i)!} \frac{(2m-n-k+i)!}{(n+k-i)!(m-n-k+i)!} \left(\sqrt{\beta_1}\right)^i \left(\sqrt{\beta_2}\right)^{n+k-i} \quad \text{for } k = 0, \dots, m-n;$$

$$\gamma_{m+k} = 2^{2(m+k)}\sum_{i=k}^{n} \frac{(2n-i)!}{i!(n-i)!} \frac{(m-k+i)!}{(m+k-i)!(i-k)!} \left(\sqrt{\beta_1}\right)^i \left(\sqrt{\beta_2}\right)^{m+k-i} \quad \text{for } k = 0, \dots, n.$$

The following result is obtained as a consequence of Theorem 3.2. It establishes that the multivariate Behrens–Fisher distribution of type 2, with odd degrees of freedom, is a finite mixture of t distributions.

Theorem 3.3. *If $\mathbf{x} = (x_1, \dots, x_k)' \sim \text{Be–Fi}_k^2(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi, 2n+1, 2m+1)$, $n, m \in \mathbb{N}$, $m \geq n$,then*

$$\mathbf{x} \sim \sum_{i=1}^{m+1} p_i\, t_k\left(\boldsymbol{\mu}, \frac{\left(\sqrt{2n+1}\cos\phi + \sqrt{2m+1}\sin\phi\right)^2}{2n+2i-1}\boldsymbol{\Sigma}, 2n+2i-1\right),$$

where the values of p_i, $i = 1, \dots, m+1$ are computed from Theorem 3.2.

The last result is important from a theoretical as well as from a practical point of view for it may provide exact results when computing, for instance, the density function of some Behrens–Fisher distributions, and may also simplify the computation of their percentiles, which are necessary to solve the multivariate Behrens–Fisher problem.

Theorem 3.4, proved in Girón et al. [16] for the univariate case, provides a representation of the multivariate Behrens–Fisher distribution as a hierarchical model which has several applications in Bayesian inference. To prove it, the application of the following lemma is necessary.

Lemma 3.1. *If $F(\lambda_i)$, $i = 1, 2$ are arbitrary distributions on $(\Lambda_i, \mathscr{B}_{\Lambda_i})$ and*

$$\mathbf{x}_1|\mathbf{x}_2 \sim \int_{\Lambda_1} \mathrm{N}_k[\mathbf{x}_1|\boldsymbol{\mu}(\lambda_1)+\mathbf{A}(\lambda_1)\mathbf{x}_2, \boldsymbol{\Sigma}(\lambda_1)]\mathrm{d}F(\lambda_1),$$

$$\mathbf{x}_2 \sim \int_{\Lambda_2} \mathrm{N}_k[\mathbf{x}_2|\mathbf{m}(\lambda_2), \mathbf{V}(\lambda_2)]\mathrm{d}F(\lambda_2),$$

then the distribution of $\mathbf{x}_1$ is expressed as

$$\int_{\Lambda_1\times\Lambda_2} \mathrm{N}_k[\mathbf{x}_1|\boldsymbol{\mu}(\lambda_1)+\mathbf{A}(\lambda_1)\mathbf{m}(\lambda_2), \boldsymbol{\Sigma}(\lambda_1)+\mathbf{A}(\lambda_1)\mathbf{V}(\lambda_2)\mathbf{A}'(\lambda_1)] \prod_{i=1}^{2} \mathrm{d}F(\lambda_i).$$

Theorem 3.4. *If*

$$\mathbf{x}|\boldsymbol{\mu} \sim \mathrm{t}_k(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}_1, \nu_1),$$
$$\boldsymbol{\mu} \sim \mathrm{t}_k(\boldsymbol{\mu}|\mathbf{0}, \boldsymbol{\Sigma}_2, \nu_2),$$

then $\mathbf{x} \sim \text{Be–Fi}_k^1\left(\mathbf{x}|\mathbf{0}, \boldsymbol{\Sigma}_1^{\frac{1}{2}}, \boldsymbol{\Sigma}_2^{\frac{1}{2}}, \nu_1, \nu_2\right)$.

Proof. It follows from the representation of the multivariate Student t distribution as a scale mixture of normals when the mixing distribution is an inverted-gamma distribution, in (3), and from Lemma 3.1, whose proof is elementary by using properties of mixtures of multivariate normal distribution. □

Theorem 3.5 provides another representation of the multivariate Behrens–Fisher distribution as a scale mixture of t distributions when the mixing distribution is an inverted beta 2 distribution.

Note that a random variable X follows an inverted beta 2 distribution with parameters m, p and n $(m, p, n > 0)$, and will be denoted as $X \sim \mathrm{Be}_2^{-1}(X|m, p, n)$ if its density function is

$$f(x|m, p, n) = \frac{\Gamma(p+m)}{\Gamma(p)\Gamma(m)} \frac{x^{m-1}n^p}{(n+x)^{p+m}}, \quad x > 0.$$

This distribution was defined in Raiffa and Schlaifer [26] and it essentially is a scale transform of a Snedecor F distribution.

Theorem 3.5. *If* $\mathbf{x} \sim \text{Be–Fi}_k^1(\mathbf{x}|\mathbf{A}_1, \mathbf{A}_2, \nu_1, \nu_2)$ *then*

$$\mathbf{x} \sim \int \mathrm{t}_k\left(\mathbf{x}\left|\mathbf{0}, \frac{\beta\nu_1+\nu_2}{\beta(\nu_1+\nu_2)}(\mathbf{A}_1\mathbf{A}_1' + \beta\mathbf{A}_2\mathbf{A}_2'), \nu_1+\nu_2\right.\right) \mathrm{dBe}_2^{-1}\left(\beta\left|\frac{\nu_1}{2}, \frac{\nu_2}{2}, \frac{\nu_2}{\nu_1}\right.\right).$$

Proof. If we set $\alpha = \lambda_1$ and $\beta = \frac{\lambda_2}{\lambda_1}$, and compute the distributions of $\alpha|\beta$ and β, then it is easily obtained

$$\alpha|\beta \sim \mathrm{Ga}^{-1}\left(\alpha\left|\frac{\nu_1+\nu_2}{2}, \frac{\nu_1\beta+\nu_2}{2\beta}\right.\right),$$
$$\beta \sim \mathrm{Be}_2^{-1}\left(\beta\left|\frac{\nu_1}{2}, \frac{\nu_2}{2}, \frac{\nu_2}{\nu_1}\right.\right).$$

Hence, the distribution of $\mathbf{x}$ can be expressed as

$$\iint \mathrm{N}_k(\mathbf{0}, \alpha(\mathbf{A}_1\mathbf{A}_1' + \beta\mathbf{A}_2\mathbf{A}_2'))\mathrm{dGa}^{-1}\left(\alpha\left|\frac{\nu_1+\nu_2}{2}, \frac{\nu_1\beta+\nu_2}{2\beta}\right.\right)\mathrm{dBe}_2^{-1}\left(\beta\left|\frac{\nu_1}{2}, \frac{\nu_2}{2}, \frac{\nu_2}{\nu_1}\right.\right),$$

and taking into account the expression (3), the theorem holds. □

4. Elliptical condition of the Behrens–Fisher distribution of type 2

The class of spherically and elliptically symmetric distributions has been widely studied in Fang, Kotz and Ng [9]. These distributions are considered an extension of the multivariate normal distribution, so they possess many properties parallel to those of this well known distribution. One of the most important is the simplification of most of the calculations involving these distributions, including their spherical or ellipsoidal credible regions, to a one dimensional integral.

In this section, the elliptical condition of the multivariate Behrens–Fisher distribution of type 2 is studied.

By considering Theorem 3.1, if $\mathbf{x} \sim \text{Be–Fi}_k^2(\mathbf{x}|\phi, \nu_1, \nu_2)$, its density function is expressed as the following non-negative function of a scalar variable $g(\mathbf{x}'\mathbf{x})$, where

$$g(y) \propto \iint (\lambda_1\cos^2\phi + \lambda_2\sin^2\phi)^{-\frac{k}{2}} \exp\left\{\frac{-y}{2(\lambda_1\cos^2\phi + \lambda_2\sin^2\phi)}\right\}\mathrm{dGa}^{-1}\left(\lambda_1\left|\frac{\nu_1}{2}, \frac{\nu_1}{2}\right.\right)\mathrm{dGa}^{-1}\left(\lambda_2\left|\frac{\nu_2}{2}, \frac{\nu_2}{2}\right.\right).$$

This result proves that the density function of $\mathbf{x}$ is constant on spheres and, as a natural consequence, its contours of equal density have spherical shape. As these statements define the spherical distributions, we conclude that the distribution of $\mathbf{x}$ is spherical.

An alternative description of a k-dimensional vector distributed according to a spherical distribution is obtained considering its stochastic representation as product of a non-negative random variate, called radial variate, and a random vector uniformly distributed on the unit sphere in $\mathbb{R}^k$. By considering this result, it is derived the equality of the distributions of the squares of the l_2-norm of this vector and its radial variate. Therefore, the square of the radial variate plays an important role in order to compute the credible regions of the vectors with spherical distributions.

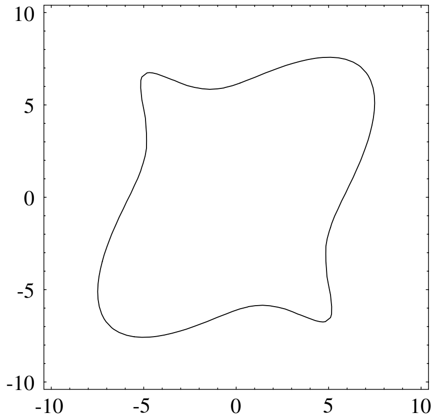


Fig. 1. Credible region of a bivariate Behrens–Fisher distribution of type 1.

For the case of the multivariate Behrens–Fisher distribution of type 2, by applying Theorems 3.1 and 3.5, the following expressions of the distribution of the square of its radial variate are derived

$$r^2 \sim \iint \mathrm{Ga}\left(r^2 \left| \frac{k}{2}, \frac{1}{2(\lambda_1 \cos^2\phi + \lambda_2 \sin^2\phi)}\right.\right) \prod_{i=1}^{2} \mathrm{dGa}^{-1}\left(\lambda_i \left| \frac{\nu_i}{2}, \frac{\nu_i}{2}\right.\right),$$

$$r^2 \sim \int \mathrm{Ga}\left(r^2 \left| \frac{k}{2}, \frac{1}{z}\right.\right) \mathrm{dCGa}^{-1}\left(z \left| \frac{\nu_1}{2}, \nu_1 \cos^2\phi, \frac{\nu_2}{2}, \nu_2 \sin^2\phi\right.\right),$$

$$r^2 \sim \int \mathrm{Be}_2^{-1}\left(\frac{k}{2}, \frac{\nu_1+\nu_2}{2}, \frac{(\beta\nu_1+\nu_2)(\cos^2\phi + \beta\sin^2\phi)}{\beta}\right) \mathrm{dBe}_2^{-1}\left(\beta \left| \frac{\nu_1}{2}, \frac{\nu_2}{2}, \frac{\nu_2}{\nu_1}\right.\right).$$

The elliptical condition of the multivariate Behrens–Fisher distribution of type 2 is obtained by considering, in a similar way as that of the multivariate normal case, that the affine transformations of a vector with a spherical distribution yields a vector distributed as an elliptical distribution. As its name indicates, the main geometric property of a vector following an elliptical distribution is the elliptical shape of its credible sets.

On the other hand, in opposition to the multivariate Behrens–Fisher distribution of type 2, the multivariate Behrens–Fisher distribution of type 1 is not generally elliptical unless the scale matrices are proportional, as it is illustrated in Fig. 1. It shows the contour, computed by Monte Carlo methods, of a credible region with probabilistic content close to 1 of a bivariate Behrens–Fisher distribution of type 1 with small values of the degrees of freedom. As we see this region is neither elliptically contoured nor even a convex set.

5. Frequentist behavior of the H.P.D. regions for the Behrens–Fisher problem

In this section, the H.P.D. regions for the multivariate Behrens–Fisher problem, where the parameter of interest is the difference of the mean vectors, and their frequentist behavior are studied.

Note that one important approach for the development of non-informative priors in Bayesian analyses is based on the *probability matching criterion* which requires matching the posterior coverage probability of a Bayesian credible set for a parameter of interest with the corresponding frequentist coverage probability asymptotically up to a certain order. An excellent monograph on this topic is due to Datta and Mukerjee [7] which provides a discussion of several probability matching criteria. In this section, the H.P.D. matching criterion is considered, demonstrating that under some suitable priors the H.P.D. regions arising from the corresponding posterior distribution have the nominal frequentist coverage.

Then, suppose that $\mathbf{X}_1 = (\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n_1})$ and $\mathbf{x}_2 = (\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n_2})$ are independent samples from the multivariate normal populations, $N_k(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N_k(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, respectively, with unknown mean vectors, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and unknown covariance matrices, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$.

Two different cases are studied depending on the form of the covariance matrices of the two normal populations. Firstly, the general case of unknown covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$; and secondly, the particular case $\boldsymbol{\Sigma}_1 = \sigma_1^2\mathbf{V}$ and $\boldsymbol{\Sigma}_2 = \sigma_2^2\mathbf{V}$, where $\mathbf{V}$ is a known matrix and σ_1^2 and σ_2^2 are unknown constants.

For the first case two priors are considered over the parameters of the problem: the commonly used independence-Jeffreys prior distribution and an alternative prior that will lead to H.P.D. regions for the difference of the mean vectors whose coverage probability matches the frequentist coverage probability more accurately that the one obtained using the independence-Jeffreys prior. For the second case, it will be shown that the behavior of the independence-Jeffreys prior, in terms of frequentist validity of the H.P.D. regions, is good and quite similar to the one obtained for the univariate case.

5.1. General case

For the first case, assuming that the parameters $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are *a priori* independent and follow the commonly used independence-Jeffreys prior distribution,

$$\pi^J(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \propto |\boldsymbol{\Sigma}_1|^{-\frac{1}{2}(k+1)}|\boldsymbol{\Sigma}_2|^{-\frac{1}{2}(k+1)},$$

it can be verified that $\boldsymbol{\mu}_1|\mathbf{X}_1$ and $\boldsymbol{\mu}_2|\mathbf{X}_2$ are *a posteriori* independent and distributed as the following Student t distributions

$$\boldsymbol{\mu}_1|\mathbf{X}_1 \sim t_k\left(\boldsymbol{\mu}_1 \,\middle|\, \bar{\mathbf{x}}_1, \frac{\mathbf{W}_1}{n_1(n_1-k)}, n_1-k\right),$$
$$\boldsymbol{\mu}_2|\mathbf{X}_2 \sim t_k\left(\boldsymbol{\mu}_2 \,\middle|\, \bar{\mathbf{x}}_2, \frac{\mathbf{W}_2}{n_2(n_2-k)}, n_2-k\right)$$

where $\bar{\mathbf{x}}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\mathbf{x}_{ji}$ and $\mathbf{W}_j = \sum_{i=1}^{n_j}(\mathbf{x}_{ji}-\bar{\mathbf{x}}_j)(\mathbf{x}_{ji}-\bar{\mathbf{x}}_j)'$ for $j = 1, 2$.

Hence, by considering (1),

$$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2|\mathbf{X}_1, \mathbf{X}_2 \sim \text{Be–Fi}_k^1\left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2, \frac{\mathbf{W}_1^{\frac{1}{2}}}{\sqrt{n_1(n_1-k)}}, \frac{\mathbf{W}_2^{\frac{1}{2}}}{\sqrt{n_2(n_2-k)}}, n_1-k, n_2-k\right). \tag{4}$$

As was shown in Section 4, the multivariate Behrens–Fisher distribution of type 1 is not elliptical; furthermore, its credible regions are not convex sets for small values of the degrees of freedom, as was illustrated in Fig. 1. For this reason, the exact analytic equations of the contours regions of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2|\mathbf{X}_1, \mathbf{X}_2$ cannot be found, so numerical or Monte Carlo methods have to be used. In Examples 1 and 2, the Monte Carlo method proposed in Chen, Shao and Ibrahim [6] is considered. It consists in the following steps: first, a sample of size n from the posterior distribution of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2|\mathbf{X}_1, \mathbf{X}_2$ is obtained; second, the density function of this difference, f, is evaluated at all the points in the sample; third, the obtained values are ordered from smaller to bigger; fourth, the value k in position αn is taken. Then, the $(1-\alpha)$ H.P.D. region for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ is $\{\mathbf{x} \in \mathbb{R}^k : f(\mathbf{x}) \geq k\}$.

The independence-Jeffreys prior is used in most of the Bayesian approaches to the multivariate Behrens–Fisher problem, see Johnson and Weerahandi [19], Nel and Groenewald [23] and Thabane and Safiul Haq [27]. However, a simulation study we have performed in Example 1, shows that it is not a H.P.D. matching prior for small values of k and sample sizes. For these cases, the frequentist coverage of the H.P.D. regions are larger than the nominal level.

Last comment suggests that another non-informative prior distribution, different of the Jeffreys prior, could be considered over the parameters of the multivariate Behrens–Fisher problem in order to result in perfect agreement between posterior probability and frequentist coverage for the H.P.D. regions for all values of k and sample sizes, including the small ones.

For this reason, we consider the following non-informative prior

$$\pi^{GC}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \propto |\boldsymbol{\Sigma}_1|^{-k}|\boldsymbol{\Sigma}_2|^{-k}.$$

Note that for $k = 1$, it is equal to the independence-Jeffreys prior. Furthermore, Geisser and Cornfield [13] used that prior distribution over the mean vector and the covariance matrix of a multivariate normal distribution and showed that it yields a posterior distribution of the mean vector which reduces to the Fisher–Cornish fiducial density.

Then, assuming that $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are *a priori* independent and follow the distribution π^{GC} defined above, it is obtained that $\boldsymbol{\mu}_1|\mathbf{X}_1$ and $\boldsymbol{\mu}_2|\mathbf{X}_2$ are *a posteriori* independent and distributed as the following Student t distributions

$$\boldsymbol{\mu}_1|\mathbf{X}_1 \sim t_k\left(\boldsymbol{\mu}_1 \,\middle|\, \bar{\mathbf{x}}_1, \frac{\mathbf{W}_1}{n_1(n_1-1)}, n_1-1\right),$$
$$\boldsymbol{\mu}_2|\mathbf{X}_2 \sim t_k\left(\boldsymbol{\mu}_2 \,\middle|\, \bar{\mathbf{x}}_2, \frac{\mathbf{W}_2}{n_2(n_2-1)}, n_2-1\right).$$

Therefore, by (1), the posterior distribution of the difference of the normal means yields

$$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2|\mathbf{X}_1, \mathbf{X}_2 \sim \text{Be–Fi}_k^1\left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2, \frac{\mathbf{W}_1^{\frac{1}{2}}}{\sqrt{n_1(n_1-1)}}, \frac{\mathbf{W}_2^{\frac{1}{2}}}{\sqrt{n_2(n_2-1)}}, n_1-1, n_2-1\right). \tag{5}$$

Using Monte Carlo methods, it can be verified that, for large values of k and sample sizes, the prior π^{GC} leads to H.P.D. regions whose coverage probabilities match the frequentist coverage probabilities as accurately as the ones obtained using the independence-Jeffreys prior. But for small values of k and sample sizes, as it occurs with other Bayesian approaches, the posterior distribution is very sensitive to the specifications of prior distribution, and the results obtained are quite different depending on the prior considered. For these values, provided that $k \neq 2$, the coverage properties are poor for the two priors, although the results obtained with π^{GC} are sensibly better than those obtained using π^J. However, for $k = 2$, the simulation results indicate excellent matching with π^{GC} even in small samples, as Example 1 shows.

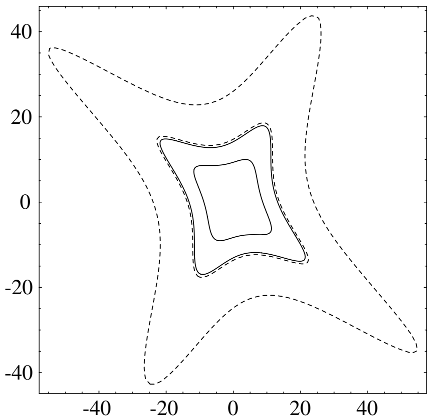


Fig. 2. 0.95 and 0.99 H.P.D. regions for the difference of the mean vectors obtained with the priors π^J (dashed lines) and π^{GC} (solid lines).

Example 1. The following samples

$$\mathbf{X}_1 = ((3.3679, -2.2717), (2.30167, 2.11408), (-6.5244, 3.8232), (-0.4193, 0.5729)),$$
$$\mathbf{X}_2 = ((2.6673, 3.5606), (-0.6602, 0.1045), (-2.4721, -5.1614), (-0.000948, -0.4111)),$$

are obtained by simulation from the populations $\mathrm{N}_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathrm{N}_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, where

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = (0, 0), \qquad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}.$$

The summary statistics are

$$\bar{\mathbf{x}}_1 = (-0.318534, 0.00260263) \qquad \mathbf{W}_1 = \begin{pmatrix} 88.468 & -56.5476 \\ -56.5476 & 36.8634 \end{pmatrix}$$
$$\bar{\mathbf{x}}_2 = (-0.116468, -0.476842) \qquad \mathbf{W}_2 = \begin{pmatrix} 20.4117 & 32.9496 \\ 32.9496 & 57.8837 \end{pmatrix}.$$

Then, assuming the independence-Jeffreys prior distribution over the parameters of the normal populations and considering (4), it is obtained that

$$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 | \mathbf{X}_1, \mathbf{X}_2 \sim \text{Be–Fi}_k^1 \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \,\middle|\, \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2, \frac{\mathbf{W}_1^{\frac{1}{2}}}{2\sqrt{2}}, \frac{\mathbf{W}_2^{\frac{1}{2}}}{2\sqrt{2}}, 2, 2 \right).$$

According to the Monte Carlo method described in this section and proposed in Chen, Shao and Ibrahim [6], and considering the representation of the multivariate Behrens–Fisher distribution given in Theorem 3.5, the $(1-\alpha)$ H.P.D. region of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 | \mathbf{X}_1, \mathbf{X}_2$ is computed for $\alpha = 0.05$ and $\alpha = 0.01$ with a sample of size $n = 10\,000$. Its equation is $\{(x, y) \in \mathbb{R}^2 : f^J(x, y) \geq k\}$ where f^J is the posterior density function of the difference of the mean vectors and the value of k depends on the level α considered; for $\alpha = 0.05$, $k = 0.0000774175$ and for $\alpha = 0.01$, $k = 3.10082 \cdot 10^{-6}$. The contours of the these regions are shown with dashed lines in Fig. 2.

Fig. 2 also shows with solid lines, the contours of the 0.95 and 0.99 H.P.D. regions of the posterior distribution of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ obtained considering π^{GC} as the prior distribution of the parameters. For this case, the distribution of the difference of the mean vectors is derived by (5) and it yields

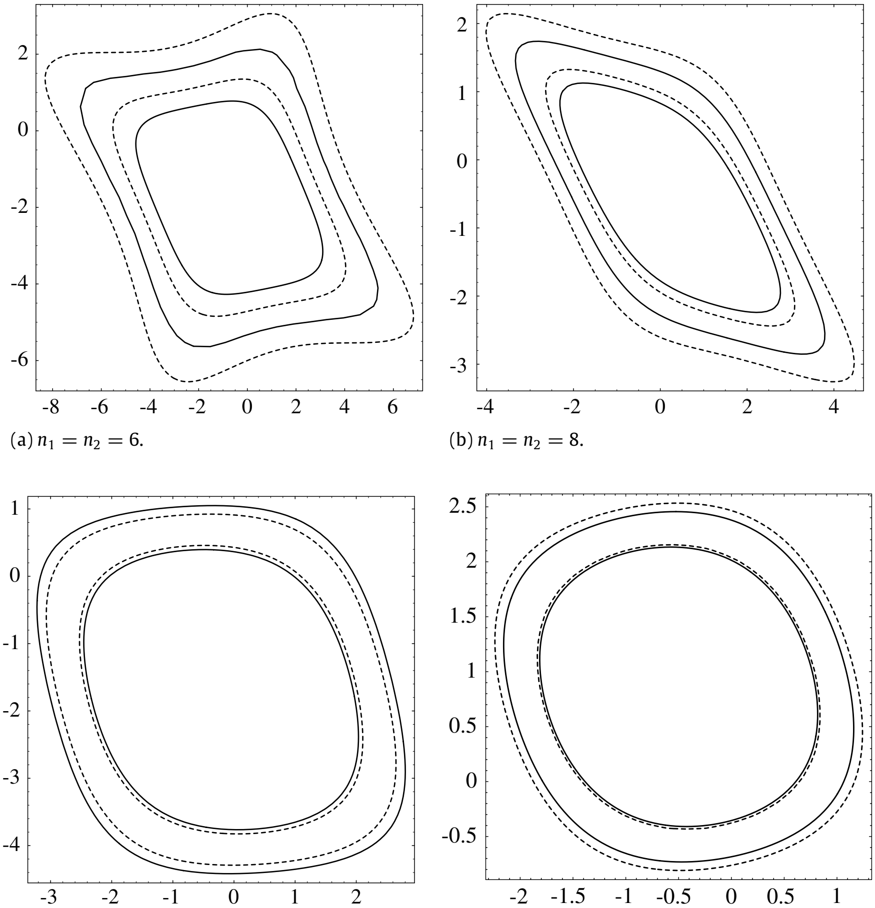
$$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 | \mathbf{X}_1, \mathbf{X}_2 \sim \text{Be–Fi}_k^1 \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \,\middle|\, \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2, \frac{\mathbf{W}_1^{\frac{1}{2}}}{2\sqrt{3}}, \frac{\mathbf{W}_2^{\frac{1}{2}}}{2\sqrt{3}}, 3, 3 \right).$$

Then, the expression of the $(1-\alpha)$ H.P.D. region of the distribution of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 | \mathbf{X}_1, \mathbf{X}_2$ is $\{(x, y) \in \mathbb{R}^2 : f^{GC}(x, y) \geq k\}$, where f^{GC} is the density function of the difference of the mean vectors, $k = 0.000212596$ for $\alpha = 0.05$ and $k = 7.84532 \cdot 10^{-6}$ for $\alpha = 0.01$.

We can also verify that the contours of the regions in Fig. 2 are not elliptical, not even convex sets, as might be expected because the posterior distribution of the difference of the mean vectors is, for the two priors, a multivariate Behrens–Fisher of type 1 and the sample sizes are small. The frequentist coverage probability of the $(1-\alpha)$ H.P.D. region of the posterior difference of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ under π^J and π^{GC} for $\alpha = 0.75, 0.5, 0.25, 0.1, 0.05$ and 0.01 is now evaluated. The computation of

Table 1
Frequentist coverage of the 0.25, 0.50, 0.75, 0.90, 0.95 and 0.99 H.P.D. regions for π^J and π^{GC}.

	$1-\alpha=0.25$	$1-\alpha=0.50$	$1-\alpha=0.75$	$1-\alpha=0.90$	$1-\alpha=0.95$	$1-\alpha=0.99$
π^J	0.390	0.698	0.908	0.981	0.995	1
π^{GC}	0.253	0.505	0.755	0.903	0.951	0.992



(a) $n_1 = n_2 = 6$. (b) $n_1 = n_2 = 8$.

(c) $n_1 = n_2 = 12$. (d) $n_1 = n_2 = 20$.

Fig. 3. 0.95 and 0.99 H.P.D. regions for the difference of the mean vectors obtained with the priors π^J (dashed lines) and π^{GC} (solid lines) for increasing values of the sample sizes.

these numerical values is based on simulation. In particular, for the values of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ at the beginning of this example, we take 10 000 independent random samples of $(\mathbf{X}_1, \mathbf{X}_2)$ from the two bivariate normal models. Under each prior, the frequentist coverage probability is estimated by the relative frequency that the $(1-\alpha)$ H.P.D. region of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 | \mathbf{X}_1, \mathbf{X}_2$ contains the value (0, 0), also computed using Monte Carlo. An inspection of Table 1 reveals that the agreement between the frequentist and posterior coverage probabilities of the H.P.D. regions is quite good for π^{GC}. However, the nominal level is exceeded for all values of α under the prior π^J.

The following example illustrates the performance of the H.P.D. regions obtained with the two priors when the sample sizes are increased.

Example 2. In this example, four samples of increasing sizes are obtained by simulation of the same multivariate normal populations of Example 1 and the 0.95 and 0.99 H.P.D. regions for the difference of the mean vectors are computed considering the two priors π^J and π^{GC}. The applied procedure is the same as the one described in Example 1.

Fig. 3 shows the contours of these H.P.D. regions. The contours with dashed lines have been obtained considering the prior π^J and the contours with solid lines have been obtained using π^{GC}. It is clear that if the sample sizes increase, the contours of the H.P.D. regions are nearly elliptical. This figure also illustrates the influence of the prior considered in the resulting posterior for small values of the sample sizes, but if these values get larger the regions obtained using π^J and π^{GC}

tend to be equal. This is also a geometric verification of the similar behavior of the two priors in terms of frequentist coverage of the H.P.D. regions for large values of the sample sizes.

5.2. Case of proportional covariance matrices

The study of the H.P.D. regions for the multivariate Behrens–Fisher problem where the parameter of interest is the difference of the mean vectors for the case of proportional covariance matrices is here considered, and their frequentist coverage is also examined.

Suppose that $\mathbf{X}_1 = (\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n_1})$ and $\mathbf{X}_2 = (\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n_2})$ are samples from the multivariate normal populations $\mathrm{N}_k(\boldsymbol{\mu}_1, \sigma_1^2\mathbf{V})$ and $\mathrm{N}_k(\boldsymbol{\mu}_2, \sigma_2^2\mathbf{V})$, respectively, where $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \sigma_1^2$ and σ_2^2 are unknown, and $\mathbf{V}$ is a known symmetric and positive definite $k \times k$ matrix.

If the parameters $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \sigma_1^2$ and σ_2^2 are assumed to be independent and follow the independence-Jeffreys prior distribution $\pi^J(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \sigma_1^2, \sigma_2^2) \propto \sigma_1^{-2}\sigma_2^{-2}$, then it is easily obtained that $\boldsymbol{\mu}_1|\mathbf{X}_1$ and $\boldsymbol{\mu}_2|\mathbf{X}_2$ are *a posteriori* independent and distributed as the following Student t distributions

$$\boldsymbol{\mu}_1|\mathbf{X}_1 \sim t_k\left(\boldsymbol{\mu}_1 \,\middle|\, \bar{\mathbf{x}}_1, \frac{s_{1V}^2}{n_1}\mathbf{V}, \nu_1\right),$$

$$\boldsymbol{\mu}_2|\mathbf{X}_2 \sim t_k\left(\boldsymbol{\mu}_2 \,\middle|\, \bar{\mathbf{x}}_2, \frac{s_{2V}^2}{n_2}\mathbf{V}, \nu_2\right),$$

where $\bar{\mathbf{x}}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\mathbf{x}_{ji}$, $s_{jV}^2 = \frac{1}{\nu_j}\sum_{i=1}^{n_j}(\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)'\mathbf{V}^{-1}(\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)$, and $\nu_j = (n_j - 1)k$ for $j = 1, 2$.

Therefore, by considering (2), the posterior distribution of the difference of the mean vectors yields

$$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2|\mathbf{X}_1, \mathbf{X}_2 \sim \text{Be–Fi}_k^2\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \,\middle|\, \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2, \sqrt{\frac{s_{1V}^2}{n_1} + \frac{s_{2V}^2}{n_2}}\,\mathbf{V}^{\frac{1}{2}}, \phi_V, \nu_1, \nu_2\right)$$

or, equivalently,

$$\left.\frac{\mathbf{V}^{-\frac{1}{2}}\{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\}}{\sqrt{\frac{s_{1V}^2}{n_1} + \frac{s_{2V}^2}{n_2}}}\right| \mathbf{X}_1, \mathbf{X}_2 \sim \text{Be–Fi}_k^2(\phi_V, \nu_1, \nu_2)$$

where ϕ_V is an angle in $[0, \frac{\pi}{2}]$ such that $\tan^2\phi_V = \frac{s_{2V}^2/n_2}{s_{1V}^2/n_1}$.

As it was shown in Section 4, the standard multivariate Behrens–Fisher distribution of type 2 is spherical; therefore, the $(1-\alpha)$ H.P.D. region of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ has an exact algebraic expression that is given by

$$\frac{((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2))'\mathbf{V}^{-1}((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2))}{\frac{s_{1V}^2}{n_1} + \frac{s_{2V}^2}{n_2}} \leq r_k^2(\phi_V, \nu_1, \nu_2; 1-\alpha) \tag{6}$$

where $r_k^2(\phi_V, \nu_1, \nu_2; 1-\alpha)$ denotes the $1-\alpha$ fractile of the distribution of the square of the radial variate of the distribution $\text{Be–Fi}_k^2(\phi_V, \nu_1, \nu_2)$. This distribution can be easily computed considering the representation given in Theorem 3.5, or Theorem 3.3 for the case of odd degrees of freedom.

A simulation study performed in Example 3 shows that the independence-Jeffreys prior distribution is a H.P.D. matching prior for this problem. Furthermore, it would be verified that the simulation results obtained in the univariate case are quite similar to those obtained for the multivariate case. This seems reasonable because the expression of the independence-Jeffreys prior distribution is the same in both cases.

Note that Ghosh and Kim [14] derived a new prior for the univariate Behrens–Fisher problem that possesses good frequentist properties, sometimes better than the one obtained using the independence-Jeffreys prior. However, that prior depends on the sample sizes and they do not suppose *a priori* independence for the variances of the two normal populations involving the univariate Behrens–Fisher problem. In spite of these comments, that prior could be considered a useful one for the particular case of the multivariate Behrens–Fisher problem studied in this subsection, but the posterior distribution of the difference of the mean vectors is not a multivariate Behrens–Fisher distribution.

Example 3. We consider the following samples

$$\mathbf{X}_1 = ((-1.9606, -0.6423), (-0.9094, 1.6935), (0.9525, 0.2567), (-0.7032, -2.9352)),$$

$$\mathbf{X}_2 = ((1.97683, 3.27627), (-3.47399, 2.12649), (2.96471, -1.34433), (3.52431, 1.44724)),$$

which have been obtained by simulation from the following normal populations $N_2(\boldsymbol{\mu}_1, \sigma_1^2\mathbf{I}_2)$ and $N_2(\boldsymbol{\mu}_2, \sigma_2^2\mathbf{I}_2)$ with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = (0, 0)$, $\sigma_1^2 = 2$ and $\sigma_2^2 = 5$.

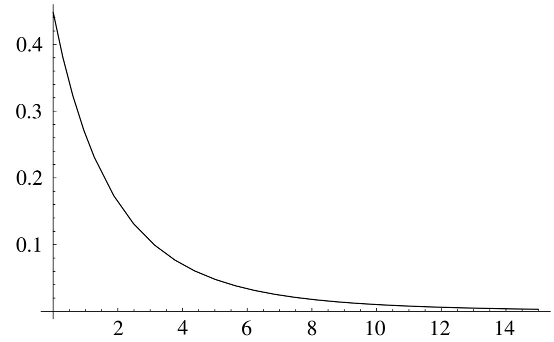


Fig. 4. Density function of the square of the radial variate of the bivariate Behrens–Fisher distribution of type 2 obtained in Example 3.

Table 2
Frequentist coverage of the 0.25, 0.50, 0.75, 0.90, 0.95 and 0.99 H.P.D. regions for π^J.

	$1-\alpha=0.25$	$1-\alpha=0.50$	$1-\alpha=0.75$	$1-\alpha=0.90$	$1-\alpha=0.95$	$1-\alpha=0.99$
π^J	0.282	0.549	0.798	0.936	0.972	0.995

The summary statistics are

$$\begin{array}{lll} \bar{\mathbf{x}}_1=(-0.655183,-0.406833) & s_1^2=4.91059 & \nu_1=6 \\ \bar{\mathbf{x}}_2=(1.24796,1.37642) & s_2^2=20.3647 & \nu_2=6 \end{array}$$

and for these values, the equation of the 0.95 H.P.D. region can be obtained by (6) where $r_2^2(\arctan^2(s_2^2/s_1^2),\nu_1,\nu_2;0.95)=9.97473$.

Fig. 4 shows, for these samples, the density function of the square of the radial variate of the distribution $\text{Be–Fi}_2^2(\arctan^2(s_2^2/s_1^2),\nu_1,\nu_2)$. This function has been computed considering the representation of the square of the radial variate given in Section 4 and derived by applying Theorem 3.5.

Table 2 shows the frequentist coverage probabilities of the H.P.D. region of the posterior difference of the mean vectors for $\alpha=0.75, 0.5, 0.25, 0.1, 0.05$ and 0.01. An inspection of it reveals that the agreement between the frequentist and posterior coverage probabilities of the H.P.D. regions is good for all values of α.

6. Discussion

The Bayesian solution of the Behrens–Fisher problem requires the computation of the posterior distribution of the difference of the mean vectors of two multivariate normal distributions when the covariance matrices are assumed different and unknown.

The form of this posterior distribution depends on the prior on all unknown parameters and, in general, it is intractable. However, for some objective priors, the posterior turns out to be the so called Behrens–Fisher distribution.

Two types of Behrens–Fisher distributions are considered in the paper and their properties and characterizations are presented in the first part of the paper.

The remainder of the paper is devoted to explore the structure of the H.P.D. regions of given probabilistic content for both types, and to examine their frequentist behavior for different objective priors using Monte Carlo methods and analytical results. The results obtained are promising and compare very well with other proposed solutions to the Behrens–Fisher problem.

Acknowledgments

We want to thank the Associate Editor and two anonymous referees for some helpful comments. This research has been supported by the research group FQM-140 and grant P07-SEJ02814 of the *Junta de Andalucìa*, and grant SEJ07-65200 of the MCYTIN.

References

[1] W.U. Behrens, Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen, Landwirtsch. Jahrb. 68 (1929) 607–837.
[2] B.M. Bennet, Note a solution of the generalized Behrens–Fisher problem, Ann. Inst. Statist. Math. 2 (1951) 87.
[3] G.E.P. Box, G.C. Tiao, Bayesian Inference in Statistical Analysis, Wiley Classics Library, New York, 1992.
[4] M.-L.G. Buot, S. Hosten, D.St.P. Richards, Counting and locating the solutions of polynomial systems of maximum likehood equations, I, J. Symbolic Comput. 41 (2006) 234–244.

[5] M.-L.G. Buot, S. Hosten, D.St.P. Richards, Counting and locating the solutions of polynomial systems of maximum likehood equations, II: the Behrens–Fisher problem, Statist. Sinica 17 (2007) 1343–1354.
[6] M.H. Chen, Q.M. Shao, J.G. Ibrahim, Monte Carlo Methods in Bayesian Computation, in: Springer Series in Statistics, 2000.
[7] G.S. Datta, R. Mukerjee, Probability Matching Priors: Higher Order Asymptotics, Springer, New York, 2004.
[8] J.M. Dickey, On a multivariate generalization of the Behrens–Fisher distribution, Ann. Math. Statist. 37 (1966) 763.
[9] K.T. Fang, S. Kotz, K.W. Ng, Symmmetric Multivariate and Related Distributions, Chapman and Hall, London, 1990.
[10] R.A. Fisher, The fiducial argument in statistical inference, Ann. Eugen. 6 (1935) 391.
[11] R.A. Fisher, The comparison of samples with possibly unequal variances, Ann. Eugen. 9 (1939) 174–180.
[12] J. Gamage, T. Mathew, S. Weerahandi, Generalized *p*-values and generalized confidence regions for themultivariate Behrens–Fisher problem and MANOVA, J. Multivariate Anal. 88 (2004) 177–189.
[13] S. Geisser, J. Cornfield, Posterior distributions for multivariate normal parameters, J. R. Stat. Soc. Ser. B 25 (1963) 368–376.
[14] J.K. Ghosh, Y.H. Kim, The Behrens–Fisher problem revisited: a Bayes-frequentist synthesis, Canad. J. Statist. 29 (2001) 5–18.
[15] F.J. Girón, C. Castillo, A note on the convolution of inverted-gamma distributions with applicationsto the Behrens–Fisher distribution, RACSAM Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. 95 (1) (2001) 39–44.
[16] F.J. Girón, M.L. Martínez, L. Imlahi, A characterization of the Behrens–Fisher distribution with applications to Bayesian inference, C. R. Acad. Sci. Paris, Sér. I 328 (1999) 701–706.
[17] A.T. James, Tests of linear hypothesis in univariate and multivariate analysis when the ratios of the population variances are unknown, Biometrika 41 (1954) 19–43.
[18] S. Johansen, The Welch–James approximation to the distribution of the residual sum of squares in a weighted linear regression, Biometrika 67 (1980) 85–92.
[19] R.A. Johnson, S. Weerahandi, A Bayesian solution to the multivariate Behrens–Fisher problem, J. Amer. Statist. Assoc. 83 (1988) 145–149.
[20] S. Kim, A practical solution to the multivariate Behrens–Fisher problem, Biometrika 79 (1992) 171–176.
[21] K. Krishnamoorthy, J. Yu, Modified Nel and van der Merwe test for the multivariate Behrens–Fisher problem, Statist. Probab. Lett. 66 (2004) 161–169.
[22] E. Moreno, B. Bertolino, W. Racugno, Default Bayesian analysis of the Behrens–Fisher problem, J. Statist. Plann. Inference 81 (1999) 323–333.
[23] D.G. Nel, P.C.N. Groenewald, A Bayesian approach to the multivariate Behrens–Fisher problem under the assumption of proportional covariance matrices, Test 2 (1993) 111–124.
[24] D.G. Nel, C.A. van der Merwe, A solution to the multivariate Behrens–Fisher problem, Comm. Statist. Theory Methods 15 (1986) 3719–3735.
[25] D.G. Nel, C.A. van der Merwe, B.K. Moser, The exact distribution of the univariate and multivariate Behrens–Fisher statistics with a comparison of several solutions in the univariate case, Comm. Statist. Theory Methods 19 (1990) 279–298.
[26] H. Raiffa, R. Schlaifer, Applied Statistical Decision Theory, John Wiley and Sons Inc., 2000.
[27] L. Thabane, M. Safiul Haq, An alternative Bayesian approach to the multivariate Behrens–Fisher problem, Comm. Statist. Simulation Comput. 28 (1) (1999) 243–258.
[28] H. Yanagihara, K.H. Yuan, Three approximate solution to the multivariate Behrens–Fisher problem, Comm. Statist. Simulation Comput. 34 (2005) 975–988.
[29] Y. Yao, An approximate degrees of freedom solution to the multivariate Behrens–Fisher problem, Biometrika 52 (1965) 139–147.