COUNTING AND LOCATING THE SOLUTIONS OF POLYNOMIAL SYSTEMS OF MAXIMUM LIKELIHOOD EQUATIONS, II: THE BEHRENS-FISHER PROBLEM

Max-Louis G. Buot and Donald St. P. Richards

Xavier University, and Penn State University and SAMSI

Abstract: Let μ be a p-dimensional vector, and let Σ_1 and Σ_2 be $p \times p$ positive definite covariance matrices. On being given random samples of sizes N_1 and N_2 from independent multivariate normal populations $N_p(\mu, \Sigma_1)$ and $N_p(\mu, \Sigma_2)$, respectively, the Behrens-Fisher problem is to solve the likelihood equations for estimating the unknown parameters μ, Σ_1, and Σ_2. It is well-known that the likelihood equations cannot be solved explicitly, and this has led to many different approaches to the Behrens-Fisher problem and with a commensurately large number of publications on the topic. We prove that for $N_1, N_2 > p$, there are, almost surely, exactly 3^p real or complex solutions of the likelihood equations. We propose a new iterative algorithm for solving the system of likelihood equations. For the case in which $p = 2$, we utilize Monte Carlo simulation to estimate how frequently a typical Behrens-Fisher is likely to have multiple real solutions; we find that multiple real solutions occur infrequently.

Key words and phrases: Behrens-Fisher problem, Bernstein's theorem, Bézout's theorem, carrier sets, facial resultant, maximum likelihood estimation, mixed volume, numerical continuation algorithms, polyhedral homotopy continuation.

1. Introduction

Suppose that $\mu \in \mathbb{R}^p$ is a constant p-dimensional vector, and suppose also that Σ_1 and Σ_2 are $p \times p$ positive definite matrices. Consider corresponding independent multivariate normal populations, $N_p(\mu, \Sigma_1)$ and $N_p(\mu, \Sigma_2)$, from which we have been given random samples $X_1, \ldots, X_{N_1}$ and $Y_1, \ldots, Y_{N_2}$, respectively. On the basis of the given data, the famous Behrens-Fisher problem is to estimate the parameters μ, Σ_1, and Σ_2 by means of the method of maximum likelihood.

It is well-known that the corresponding system of likelihood equations cannot be solved explicitly, and that has led many to propose alternative solutions to the Behrens-Fisher problem. As a consequence, the literature on the Behrens-

Fisher problem is substantial, reflecting the intense interest which the problem has generated since its inception, and indeed the problem has generated both an extensive philosophical discussion in addition to efforts to derive solutions which are optimal for statistical inference. In this paper, we study the problem of determining the number of solutions of the likelihood equations.

For the case in which $p = 1$, i.e., the univariate Behrens-Fisher problem, there are three unknown scalar parameters, *viz.*, μ, the common mean, and σ_1^2 and σ_2^2, the population variances. For this problem, Sugiura and Gupta (1987) reduced the system of equations to a cubic equation in μ and deduced that, almost surely, there are three complex solutions. They observed also that the likelihood equation tended to have multiple solutions if σ_1^2 and σ_2^2 are small in comparison with μ, and otherwise that the likelihood equation usually has a unique solution. In the sequel, we shall provide an explanation of these phenomena in the general multivariate case.

Drton (2006) has also studied the univariate Behrens-Fisher problem and observed that the system of three equations for μ, σ_1^2, and σ_2^2 has the algebraic structure of a *lexicographic Gröbner basis* (Pachter and Sturmfels (2005), p. 86). By analyzing the discriminant of this system of equations, Drton shows that the paucity of three real solutions is a consequence of the Gröbner basis formulation of the likelihood equations. Therefore, it is natural to conjecture that similar results hold in the general multivariate setting.

In this paper, as in our earlier article (Buot and Richards, 2006), we apply results from the theory of algebraic geometry to study the solution set of the system of likelihood equations for the multivariate Behrens-Fisher problem. Remarkably, it turns out that the exact number of solutions of the likelihood equations can be determined. Generalizing the univariate result described earlier, we shall prove the following result:

Theorem. Suppose that $N_1, N_2 > p$. Then, almost surely, there are exactly 3^p real or complex solutions of the system of likelihood equations for the multivariate Behrens-Fisher problem. In particular, almost surely, there always exists at least one real solution.

We remark that, by an appeal to algebraic geometry, the number of solutions can be shown to be finite, almost surely. Indeed, by Huber and Sturmfels (1995), a system of p equations in p variables has a finite number of solutions if all p "facial resultants" of the system are non-zero; here, a facial resultant is a certain determinant in the coefficients of the system of equations. In our case, the entries of these determinants are polynomials in the continuous random variables $\{X_j\}$ and $\{Y_j\}$, so we find that the facial resultants themselves are continuous random variables. Therefore, almost surely, all p facial resultants are non-zero, from which we deduce that the number of solutions of the system is finite. Thus, the significance of the theorem stems from the formula for the exact number of solutions.

2. Remarks on the likelihood equations

Denote by $\bar{X}$ and $\bar{Y}$ be the sample means of the samples from $N_p(\mu, \Sigma_1)$ and $N_p(\mu, \Sigma_2)$, respectively. By standard calculations (cf., Mardia, *et al.* (1979), p. 142), we find that the likelihood equations for estimating μ, Σ_1 and Σ_2 are:

$$\begin{aligned}\widehat{\Sigma}_1 &= N_1^{-1} \sum_{j=1}^{N_1} (X_j - \widehat{\mu})(X_j - \widehat{\mu})', \\ \widehat{\Sigma}_2 &= N_2^{-1} \sum_{j=1}^{N_1} (Y_j - \widehat{\mu})(Y_j - \widehat{\mu})'\end{aligned} \tag{2.1}$$

and

$$(N_1 \widehat{\Sigma}_1^{-1} + N_2 \widehat{\Sigma}_2^{-1}) \widehat{\mu} = N_1 \widehat{\Sigma}_1^{-1} \bar{X} + N_2 \widehat{\Sigma}_2^{-1} \bar{Y}. \tag{2.2}$$

Some authors have proposed the following iterative algorithm for solving (2.1) and (2.2):

(1) Begin the iteration with initial estimates $\widehat{\Sigma}_{i,0} = \widetilde{S}_i$, $i = 1, 2$, where

$$\begin{aligned}\widetilde{S}_1 &= N_1^{-1} \sum_{j=1}^{N_1} (X_j - \bar{X})(X_j - \bar{X})', \\ \widetilde{S}_2 &= N_2^{-1} \sum_{j=1}^{N_2} (Y_j - \bar{Y})(Y_j - \bar{Y})'.\end{aligned} \tag{2.3}$$

(2) Apply (2.2) to calculate $\widehat{\mu}_0$, the corresponding estimate of μ, in the form

$$\widehat{\mu}_0 = (N_1 \widehat{\Sigma}_{1,0}^{-1} + N_2 \widehat{\Sigma}_{2,0}^{-1})^{-1} N_1 \widehat{\Sigma}_{1,0}^{-1} \bar{X} + N_2 \widehat{\Sigma}_{2,0}^{-1} \bar{Y}.$$

(3) Using the value of $\widehat{\mu}_0$ obtained in Step 2, calculate $\widehat{\Sigma}_{i,1}$, an updated value of $\widehat{\Sigma}_{i,0}$ using the formulas,

$$\widehat{\Sigma}_{1,1} = \widetilde{S}_1 + (\bar{X} - \widehat{\mu}_0)(\bar{X} - \widehat{\mu}_0)', \quad \widehat{\Sigma}_{2,1} = \widetilde{S}_2 + (\bar{Y} - \widehat{\mu}_0)(\bar{Y} - \widehat{\mu}_0)',$$

which are a consequence of (2.1).

(4) Return to Step 2 and update $\widehat{\mu}_j$ until the sequences $\widehat{\Sigma}_{1,j}$ and $\widehat{\Sigma}_{2,j}$, $j = 1, 2, 3, \ldots$, converge.

We have not been able to locate in the literature an investigation of the convergence properties of this algorithm. We note that the algorithm rests on the implicit assumptions that the initial values of the $\widehat{\Sigma}_i$ will lead to convergence and that the objective function has a unique local – and hence global – maximum. It is precisely this issue which is at the heart of any attempt to solve (2.1) and (2.2). If the likelihood function were found to be multimodal, a phenomenon which has been encountered recently by Drton and Richardson (2004) in a study of seemingly unrelated regression models, then any numerical algorithm for solving the system of likelihood equations necessarily must include some information about the choice of initial values.

At first glance, it appears that the likelihood equations are a system of $p(p+2)$ equations in $p(p+2)$ variables comprising the p components of μ and the $p(p+1)/2$ independent entries of both Σ_1 and Σ_2. However, a closer inspection of (2.1) and (2.2) reveals that if $\widehat{\mu}$ is known then $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$ are determined completely. This observation raises the possibility that the system of $p(p+2)$ equations may be reducible to p equations and, in fact, we shall accomplish such a reduction. By using the special structure of (2.1) and by applying some judicious algebraic manipulations, we shall eliminate $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$ from (2.2), thereby obtaining a system of p third-degree equations in the variables $\widehat{\mu}_1, \ldots, \widehat{\mu}_p$.

For a given random sample from each normal population, once we obtain a numerical solution $\widehat{\mu}$ to (2.1) and (2.2), the resulting numerical values of $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$ in (2.1), being sums of rank-one positive semidefinite matrices, are also positive semidefinite. It then follows from (2.1) that, for $N_1, N_2 > p$, each real solution $\widehat{\mu}$ automatically results in numerical solutions for Σ_1 and Σ_2 which are positive definite, almost surely. Hence, in applying algebraic geometric methods, all real solutions for $\widehat{\mu}$ are permissible; i.e., in the case of real values of $\widehat{\mu}$, there is no need to carry out saturation, the process of discarding solutions which lead to parameter estimates not in conformity with restrictions basic to the underlying statistical problem.

3. The proof of the theorem

We apply to the sums in (2.1) the standard procedure of writing each term $X_i - \widehat{\mu}$ as $X_i - \bar{X} + \bar{X} - \widehat{\mu}$, and similarly for each term $Y_i - \widehat{\mu}$. This leads to the formulas

$$\widehat{\Sigma}_1 = \widetilde{S}_1 + (\bar{X} - \widehat{\mu})(\bar{X} - \widehat{\mu})' \tag{3.1}$$

and

$$\widehat{\Sigma}_2 = \widetilde{S}_2 + (\bar{Y} - \widehat{\mu})(\bar{Y} - \widehat{\mu})' \tag{3.2}$$

where $\widetilde{S}_1$ and $\widetilde{S}_2$ are defined in (2.3). By a special case of Woodbury's theorem (cf., Muirhead (1982), p. 580, Theorem A5.1) we have, for any nonsingular $p \times p$ matrix M and any column vector $v \in \mathbb{R}^p$,

$$(M + vv')^{-1} = M^{-1} - \frac{M^{-1}vv'M^{-1}}{1 + v'M^{-1}v}.$$

Multiplying the latter equation on each side from the right by v and simplifying the result, we obtain

$$\begin{aligned} (M + vv')^{-1}v &= M^{-1}v - \frac{M^{-1}vv'M^{-1}v}{1 + v'M^{-1}v} \\ &= \frac{(1 + v'M^{-1}v)M^{-1}v - (M^{-1}v)(v'M^{-1}v)}{1 + v'M^{-1}v} \\ &= \frac{M^{-1}v}{1 + v'M^{-1}v}. \end{aligned}$$

Setting $M = \widetilde{S}_1$ and $v = \bar{X} - \widehat{\mu}$, we obtain

$$\begin{aligned} \widehat{\Sigma}_1^{-1}(\bar{X} - \widehat{\mu}) &\equiv \big(\widetilde{S}_1 + (\bar{X} - \widehat{\mu})(\bar{X} - \widehat{\mu})'\big)^{-1}(\bar{X} - \widehat{\mu}) \\ &= \frac{\widetilde{S}_1^{-1}(\bar{X} - \widehat{\mu})}{1 + (\bar{X} - \widehat{\mu})'\widetilde{S}_1^{-1}(\bar{X} - \widehat{\mu})}, \end{aligned} \tag{3.3}$$

and, similarly,

$$\widehat{\Sigma}_2^{-1}(\bar{Y} - \widehat{\mu}) = \frac{\widetilde{S}_2^{-1}(\bar{Y} - \widehat{\mu})}{1 + (\bar{Y} - \widehat{\mu})'\widetilde{S}_2^{-1}(\bar{Y} - \widehat{\mu})}. \tag{3.4}$$

On rewriting (2.2) as

$$N_1\widehat{\Sigma}_1^{-1}(\bar{X}-\widehat{\mu})+N_2\widehat{\Sigma}_2^{-1}(\bar{Y}-\widehat{\mu})=0,$$

it follows from (3.3) and (3.4) that (2.2) is equivalent to

$$\frac{N_1\widetilde{S}_1^{-1}(\bar{X}-\widehat{\mu})}{1+(\bar{X}-\widehat{\mu})'\widetilde{S}_1^{-1}(\bar{X}-\widehat{\mu})}+\frac{N_2\widetilde{S}_2^{-1}(\bar{Y}-\widehat{\mu})}{1+(\bar{Y}-\widehat{\mu})'\widetilde{S}_2^{-1}(\bar{Y}-\widehat{\mu})}=0.$$

For $\widehat{\mu}\in\mathbb{R}^p$, the denominators in the latter equation are strictly positive so, by clearing denominators, we obtain

$$\begin{aligned}N_1\big(1+(\bar{Y}-\widehat{\mu})'\widetilde{S}_2^{-1}(\bar{Y}-\widehat{\mu})\big)\widetilde{S}_1^{-1}(\bar{X}-\widehat{\mu})&\\+N_2\big(1+(\bar{X}-\widehat{\mu})'\widetilde{S}_1^{-1}(\bar{X}-\widehat{\mu})\big)\widetilde{S}_2^{-1}(\bar{Y}-\widehat{\mu})&=0,\end{aligned}\tag{3.5}$$

a system of polynomial equations for $\widehat{\mu}_1,\ldots,\widehat{\mu}_p$. By inspection, we see that each of these equations is of degree 3. Therefore, by Bézout's theorem (cf., Cox, *et al.* (1998), Sturmfels (2002)), the system of equations (3.5) has at most 3^p real or complex solutions.

To derive the exact number of solutions of the system, we now apply the results of Bernstein (1975), Khovanskiĭ(1978), and Kushnirenko (1976), now commonly known as BKK theory; cf., Cox, *et al.* (1998), Sturmfels (2002). Our use of BKK theory follows on our earlier application (Buot and Richards, 2006) in work on mixture models.

Thus, let $e_1,\ldots,e_p$ denote the standard basis for the vector space $\mathbb{R}^p$. In each of the p equations in (3.5), the carrier set of the quadratic terms $1+(\bar{Y}-\widehat{\mu})'\widetilde{S}_2^{-1}(\bar{Y}-\widehat{\mu}$ and $1+(\bar{X}-\widehat{\mu})'\widetilde{S}_1^{-1}(\bar{X}-\widehat{\mu})$ is the Minkowski sum, $\{0,e_1,\ldots,e_p\}+\{0,e_1,\ldots,e_p\}$. Further, the linear terms arising from the products $\widetilde{S}_1^{-1}(\bar{X}-\widehat{\mu})$ and $\widetilde{S}_2^{-1}(\bar{Y}-\widehat{\mu})$, have carrier set $\{0,e_1,\ldots,e_p\}$. Therefore, each of the p equations in (3.5) has common carrier set

$$\Pi:=\{0,e_1,\ldots,e_p\}+\{0,e_1,\ldots,e_p\}+\{0,e_1,\ldots,e_p\}.\tag{3.6}$$

Letting $\mathbb{N}_0$ denote the set of nonnegative integers, it follows that

$$\Pi=\big\{\sum_{j=1}^p n_je_j : n_j\in\mathbb{N}_0, 1\le j\le p, \sum_{j=1}^p n_j\le 3\big\}.\tag{3.7}$$

Indeed, each n_j in (3.7) is simply the number of times that e_j is chosen in forming the elements of the Minkowski sum (3.6).

It now follows from (3.7) that $\mathcal{C}(\Pi)$, the convex hull of Π, is

$$\begin{aligned}\mathcal{C}(\Pi) &= \big\{\sum_{j=1}^{p} t_j e_j : t_j \ge 0, 1 \le j \le p, \sum_{j=1}^{p} t_j \le 3\big\} \\ &\equiv \big\{(t_1, \ldots, t_p) : t_j \ge 0, 1 \le j \le p, \sum_{j=1}^{p} t_j \le 3\big\}.\end{aligned}$$

By the well-known Dirichlet integral, the volume of $\mathcal{C}(\Pi)$ is $3^p/\Gamma(p+1) = 3^p/p!$. By BKK theory (Kushnirenko, 1976), it follows that the number of isolated solutions of (3.5) in $(\mathbb{C}^*)^p$ is $p! \cdot \mathrm{Vol}(\mathcal{C}(\Pi)) = 3^p$, where $\mathbb{C}^*$ denotes the set of non-zero complex numbers.

Because the data $\{X_j\}$ and $\{Y_j\}$ are normally distributed, then the solutions of (3.5) are continuous random variables. Therefore, it follows that all solutions are isolated, almost surely. Moreover, since the number of solutions is odd then, due to the fact that solutions appear in complex conjugate pairs, there exists at least one real solution. This completes the proof of the theorem. $\square$

Remark. Denoting by $L(\mu, \Sigma_1, \Sigma_2)$ the likelihood function for the Behrens-Fisher problem, it is straightforward to show that

$$L(\widehat{\mu}, \widehat{\Sigma}_1, \widehat{\Sigma}_2) = (2\pi e)^{-(N_1+N_2)p/2}\, |\widehat{\Sigma}_1|^{-N_1/2}\, |\widehat{\Sigma}_2|^{-N_2/2}.$$

By (3.1) and (3.2), we obtain

$$|\widehat{\Sigma}_1| = |\widetilde{S}_1| \cdot \big(1 + (\bar{X} - \widehat{\mu})' \widetilde{S}_1^{-1} (\bar{X} - \widehat{\mu})\big)$$

and

$$|\widehat{\Sigma}_2| = |\widetilde{S}_2| \cdot \big(1 + (\bar{Y} - \widehat{\mu})' \widetilde{S}_2^{-1} (\bar{Y} - \widehat{\mu})\big).$$

Therefore

$$\begin{aligned}L(\widehat{\mu}, \widehat{\Sigma}_1, \widehat{\Sigma}_2) = (2\pi e)^{-(N_1+N_2)p/2}\, |\widetilde{S}_1|^{-N_1/2}\, |\widetilde{S}_2|^{-N_2/2} \\ \times \big(1 + (\bar{X} - \widehat{\mu})' \widetilde{S}_1^{-1} (\bar{X} - \widehat{\mu})\big)^{-N_1/2} \big(1 + (\bar{Y} - \widehat{\mu})' \widetilde{S}_2^{-1} (\bar{Y} - \widehat{\mu})\big)^{-N_2/2}.\end{aligned}$$

It is clear now that, to find the maximum value of L, we need to minimize

$$\big(1 + (\bar{X} - \widehat{\mu})' \widetilde{S}_1^{-1} (\bar{X} - \widehat{\mu})\big)^{N_1/2} \big(1 + (\bar{Y} - \widehat{\mu})' \widetilde{S}_2^{-1} (\bar{Y} - \widehat{\mu})\big)^{N_2/2}. \tag{3.8}$$

While this expression can be easily evaluated computationally at each real solution of the score equation, it would be good to have further information about an algebraic expression for the minimal value of (3.8).

The expression (3.8) also provides us with information about the existence of multiple solutions of the likelihood equations. If $\bar{X} = \bar{Y}$ then it follows from (3.8) that the unique solution of the likelihood equations is $\widehat{\mu} = \bar{X} = \bar{Y}$. More generally, the smaller the distance between $\bar{X}$ and $\bar{Y}$ (in the sense that $(\bar{X} - \widehat{\mu})'\widetilde{S}_1^{-1}(\bar{X} - \widehat{\mu})$ and $(\bar{Y} - \widehat{\mu})'\widetilde{S}_2^{-1}(\bar{Y} - \widehat{\mu})$ are close in value), the more likely it is that the likelihood equations will have a unique real solution. We see then that Behrens-Fisher problems encountered in data analysis are likely to have a unique solution, for inference about the difference between the means of multivariate normal populations would be carried out by a practical investigator only if the difference between population means, *viz.*, μ, is small.

The above observations also lead us to propose a new iterative algorithm for solving the likelihood equations (3.1) and (3.2). For any vectors $u, v \in \mathbb{R}^p$, it is a simple algebraic identity that

$$u'u - v'v = (u - v)'(u + v).$$

Applying this identity with $u = \widetilde{S}_1^{-1/2}(\bar{X} - \widehat{\mu})$ and $v = \widetilde{S}_2^{-1/2}(\bar{Y} - \widehat{\mu})$ and simplifying the resulting expressions, we obtain

$$\begin{aligned}
&(\bar{X} - \widehat{\mu})'\widetilde{S}_1^{-1}(\bar{X} - \widehat{\mu}) - (\bar{Y} - \widehat{\mu})'\widetilde{S}_2^{-1}(\bar{Y} - \widehat{\mu}) \\
&\qquad = \big(\widetilde{S}_1^{-1/2}\bar{X} - \widetilde{S}_2^{-1/2}\bar{Y} - (\widetilde{S}_1^{-1/2}\widehat{\mu} - \widetilde{S}_2^{-1/2})\widehat{\mu}\big)' \\
&\qquad\qquad \times \big(\widetilde{S}_1^{-1/2}\bar{X} + \widetilde{S}_2^{-1/2}\bar{Y} - (\widetilde{S}_1^{-1/2} + \widetilde{S}_2^{-1/2})\widehat{\mu}\big).
\end{aligned}$$

To ensure that $(\bar{X} - \widehat{\mu})'\widetilde{S}_1^{-1}(\bar{X} - \widehat{\mu})$ and $(\bar{Y} - \widehat{\mu})'\widetilde{S}_2^{-1}(\bar{Y} - \widehat{\mu})$ are close in value, we choose $\widehat{\mu}$ so that the term $\widetilde{S}_1^{-1/2}\bar{X} + \widetilde{S}_2^{-1/2}\bar{Y} - (\widetilde{S}_1^{-1/2} + \widetilde{S}_2^{-1/2})\widehat{\mu}$ is small.

We then propose a new iterative algorithm for solving the likelihood equations for the Behrens-Fisher problem, as follows:

(1) Start with the initial value for $\widehat{\mu}$:

$$\widehat{\mu}_0 = (\widetilde{S}_1^{-1/2} + \widetilde{S}_2^{-1/2})^{-1}(\widetilde{S}_1^{-1/2}\bar{X} + \widetilde{S}_2^{-1/2}\bar{Y})$$

(2) Using (3.1) and (3.2), calculate corresponding estimates of $\widehat{\Sigma}_i$, $i = 1, 2$:

$$\begin{aligned}\widehat{\Sigma}_{1,0} &= \widetilde{S}_1 + (\bar{X} - \widehat{\mu}_0)(\bar{X} - \widehat{\mu}_0)' \\ \widehat{\Sigma}_{2,0} &= \widetilde{S}_2 + (\bar{Y} - \widehat{\mu}_0)(\bar{Y} - \widehat{\mu}_0)'\end{aligned}$$

(3) Update $\hat{\mu}_0$ using the likelihood equation (2.2):

$$\widehat{\mu}_1 = (N_1\widehat{\Sigma}_{1,0}^{-1} + N_2\widehat{\Sigma}_{2,0}^{-1})^{-1}\big(N_1\widehat{\Sigma}_{1,0}^{-1}\bar{X} + N_2\widehat{\Sigma}_{2,0}^{-1}\bar{Y}\big)$$

(4) Return to Step 2 and repeat the process until both $\hat{\Sigma}_{1,j}$ and $\hat{\Sigma}_{2,j}$ converge

As with the earlier algorithm at (2.3), it is a challenging problem to ascertain the convergence properties of this algorithm.

4. Simulations for bivariate Behrens-Fisher problems

Having determined the number of solutions of the system of likelihood equations (3.5) it is natural to seek the number of *real* solutions, for it is those solutions which are of interest in statistical inference. Not surprisingly, it appears to be difficult to determine an algebraic expression for the number of real solutions of the system; indeed, this is also the case for the general theory of systems of polynomial equations.

To study the real solutions of the system (3.5), we consider the case in which $p = 2$, presenting empirical evidence that multiple solutions occur rarely. In these computations, each simulation run consists of samples from two bivariate normal distributions, $N_2(\mu, \Sigma_1)$ and $N_2(\mu, \Sigma_2)$, in which the sample sizes N_1 and N_2, and the parameters μ, Σ_1, and Σ_2, are randomly generated. The solutions of the resulting likelihood equations (3.5) were computed numerically using `PHCpack` (Verschelde, 1999), a software package which implements polyhedral homotopy continuation methods for solving systems of polynomial equations.

The results of our simulations show that multiple solutions can occur. For example, for $N_1 = 11$, $N_2 = 5$, and the summary statistics

$$\bar{X} = \begin{pmatrix} -1.5516 \\ -9.4713 \end{pmatrix}, \qquad \widetilde{S}_1 = \begin{pmatrix} 0.3998 & -0.1026 \\ -0.1026 & 0.2378 \end{pmatrix},$$

$$\bar{Y} = \begin{pmatrix} -1.9175 \\ -10.4805 \end{pmatrix}, \qquad \widetilde{S}_2 = \begin{pmatrix} 0.4193 & 0.0792 \\ 0.0792 & 0.0334 \end{pmatrix},$$

the real solutions for μ are

$$\begin{pmatrix} -1.3570 \\ -10.2957 \end{pmatrix}, \quad \begin{pmatrix} -1.2478 \\ -9.9902 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} -1.4451 \\ -9.6333 \end{pmatrix}.$$

This example seems, however, to be a rare exception. In fact, our simulations found that the bivariate Behrens-Fisher likelihood equations (3.5) had one real solution about 99.5% of the time, three real solutions about 0.5% of the time, and we found no instances in which the equations had five or more real solutions.

To test for distinctions between the case of small and large samples in the bivariate case, we performed simulations in which N_1 and N_2 were randomly generated between 3 and 15. The outcomes are given as follows:

Table 4.1: Simulations with $3 \le N_1, N_2 \le 15$

Number of solutions	Frequency	Percentage
1	4450	99.29%
3	32	0.71%

As noted above, none of these simulation resulted in more than three real solutions.

In the case of larger samples, our simulations resulted in the following outcomes:

Table 4.2: Simulations with $15 \le N_1, N_2 \le 60$

Number of solutions	Frequency	Percentage
1	4404	99.46%
3	24	0.54%

Here again, no simulation resulted in more than three real solutions.

In summary, there seems to be little chance that a randomly generated, two-dimensional Behrens-Fisher problem will have three or more real solutions, and there is a high chance that it will have a unique real solution.

Acknowledgment. We thank Mathias Drton and Bernd Sturmfels for discussions on early drafts of this manuscript.

References

Bernstein, D. N. (1975). The number of roots of a system of equations. *Functional. Anal. Appl.*, **9**, 1–14.

Buot, M.-L. G. and Richards, D. St. P. (2006). Counting and locating the solutions of polynomial systems of maximum likelihood equations, I. *J. Symbolic Comput.* **41**, 234–244.

Cox, D. A., Little, J., and O'Shea, D. (1998). *Using Algebraic Geometry.* Springer, New York.

Drton, M. and Richardson, T. (2004). Multimodality of the likelihood in the bivariate seemingly unrelated regressions model. *Biometrika* **91**, 383–392.

Drton, M. (2006). Private communication on the univariate Behrens-Fisher problem.

Huber, B. and Sturmfels, B. (1995). A polyhedral method for solving sparse polynomial systems. *Math. Comp.*, **64**, 1541–1555.

Khovanskiĭ, A. G. (1978). Newton polyhedra, and the genus of complete intersections. *Funct. Anal. Appl.*, **12**, 51–61.

Kushnirenko, A. G. (1976). Newton polytopes and the Bézout theorem. *Funct. Anal. Appl.*, **10**, 233–235.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis.* Academic Press, New York.

Pachter, L. and Sturmfels, B. (2005). *Algebraic Statistics for Computational Biology.* Cambridge University Press, Cambridge, U.K.

Sturmfels, B. (2002). *Solving Systems of Polynomial Equations.* NSF-CBMS Regional Conference Series in Mathematics, Vol. **97**. Amer. Math. Soc., Providence, RI.

Sugiura, N. and Gupta, A. K. (1987). Maximum likelihood estimates for the Behrens-Fisher problem. *J. Japan Statist. Soc.*, **17**, 55–60.

Verschelde, J. (1999). Algorithm 795: PHCpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Trans. Math. Software*, **25**, 251–276.

Department of Mathematics and Computer Science, Xavier University, Cincinnati, OH 45207.

E-mail: buotm@xavier.edu

Department of Statistics, Penn State University, University Park, PA 16802, and the Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC 27709.

E-mail: richards@stat.psu.edu