Counting and Locating the Solutions of Polynomial Systems of Maximum Likelihood Equations, I

Max-Louis G. Buot [a] and Donald St. P. Richards [b]

[a] *Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213 and National Institute of Statistical Sciences, Research Triangle Park, NC 27709*

[b] *Department of Statistics, Penn State University, University Park, PA 16802*

Abstract

In statistics, mixture models consisting of several component subpopulations are used widely to model data drawn from heterogeneous sources. In this paper, we consider maximum likelihood estimation for mixture models in which the only unknown parameters are the component proportions. By applying the theory of multivariable polynomial equations, we derive bounds for the number of isolated roots of the corresponding system of likelihood equations. If the component densities belong to certain familiar continuous exponential families, including the multivariate normal or gamma distributions, then our upper bound is, almost surely, the exact number of solutions. To illustrate our main result, we apply homotopy continuation methods to the maximum likelihood equations of a three-component mixture model based on simulated data.

Key words: Bernstein's theorem, carrier sets, EM algorithm, facial resultant, finite mixture model, genetic algorithms, homotopy continuation methods, maximum likelihood estimation, mixed volume, numerical continuation algorithms.

1 Introduction

This article is the first of several in which we shall study maximum likelihood equations which are of, or are reducible to, polynomial type. Polynomial systems of likelihood equations arise in many aspects of statistical inference, including finite mixture models (Lindsay, 1995; Titterington, et al., 1985), categorical data analysis (Fienberg and Meyer, 1983; Diaconis and Sturmfels, 1998); Behrens-Fisher problems and factor analysis (Anderson, 2003); inference subject to missing data (Little and Rubin, 2002); covariance matrix esti-

Preprint submitted to Elsevier Science *8 June 2004*

mation under linear constraints (Anderson, 1973); seemingly unrelated regressions (Drton and Richardson, 2004); and order restricted hypothesis testing (Hoferkamp and Peddada, 2002). In subsequent articles, we shall investigate polynomial equations arising in these and other areas.

In this research program, our primary goals are threefold. First, we shall determine conditions under which polynomial systems of maximum likelihood equations have a finite number of solutions. Moreover, in cases in which the number of solutions is finite, we shall derive bounds or exact formulas for the number of solutions. Our approach is motivated by recent articles of Sturmfels (1998, 2000, 2002) on solving systems of polynomial equations.

Our second goal deals with explicit formulas for the roots of these equations. The calculation of maximum likelihood estimates is generally a nontrivial optimization problem, hence the corresponding equations usually are solved by means of iterative methods, e.g., Newton's method, EM algorithms, genetic algorithms. We will indicate explicit solutions in the form of infinite series, with the goal of applying these explicit formulas to study the exact, small sample, distributional properties of the maximum likelihood estimators.

In general, classical iterative schemes require careful choice of initial values to ensure convergence to global maxima. Hence, our third goal is to apply iterative algorithms, called *homotopy*, or *numerical continuation methods*, to solve our systems of equations. Continuation methods are advantageous not only in being impervious to the choice of initial values, but also because they locate *all* solutions to the system of polynomial equations. We note that the application of continuation methods appears to be new to the statistics literature.

In this paper, we consider the estimation of component weights in finite mixture models in which the component density functions are specified completely. This problem has been studied widely (Lindsay, 1995; Titterington, et al., 1985) with numerical methods as the general approach to computing parameter estimates, and it serves as a test case for the application of polynomial methods. We shall apply the theory of polynomial equations to deduce additional information about the system of maximum likelihood equations.

In Section 2, we introduce the general mixture model and related statistical notation. In Section 3, which is devoted to the case of two components, we derive the maximum likelihood equation, some properties of the corresponding solution, and indicate an infinite series formula for the estimator. In Section 4, we derive the system of likelihood equations for the general mixture model, reduce the system to polynomial form, and derive an upper bound or exact formula for the number of nonzero complex roots of the system. In Section 5, we provide an example in which homotopy continuation methods are applied to estimate the weights in a three-component mixture model.

2 Finite mixture models

Consider a random vector $X \in \mathbb{R}^d$ with probability density function

$$f(x) = \sum_{j=1}^{m} \pi_j f_j(x), \tag{2.1}$$

$x \in \mathbb{R}^d$, where $f_1, \ldots, f_m$ are density functions on $\mathbb{R}^d$ and $(\pi_1, \ldots, \pi_m)$ belongs to the simplex $\mathcal{S}_m = \{(\pi_1, \ldots, \pi_m) : \pi_j \geq 0, j = 1, \ldots, m; \pi_1 + \cdots + \pi_m = 1\}$. Mixture models such as (2.1) arise naturally when random samples are drawn from a population consisting of homogeneous sub-populations. The model (2.1) is a *finite mixture model* for X, the parameters $\pi_1, \ldots, \pi_m$ are the *mixing proportions* or *component weights*, and $f_1, \ldots, f_m$ are the *component density functions*. We assume that m is known and that $f_1, \ldots, f_m$ are completely specified, so the estimation of $\pi_1, \ldots, \pi_m$ is the problem of interest. Mixture models with known component densities have been studied extensively (see Lindsay, 1995; Titterington, et al., 1985) and have provided much insight into general mixture models in which not all densities f_j are fully specified.

Let $X_1, \ldots, X_n$ be a random sample drawn from X. The likelihood function,

$$L(\pi_1, \ldots, \pi_m) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \Big[\sum_{j=1}^{m} \pi_j f_j(x_i)\Big], \tag{2.2}$$

is the joint density of $X_1, \ldots, X_n$, regarded as a function of the parameters $\pi_1, \ldots, \pi_m$. Since $\pi_m = 1 - (\pi_1 + \cdots + \pi_{m-1})$ then we may view L as a function of $\pi_1, \ldots, \pi_{m-1}$. A maximum likelihood estimate $(\hat{\pi}_1, \ldots, \hat{\pi}_m)$ of $(\pi_1, \ldots, \pi_m)$ is a value at which the function L is maximized. For the case in which the f_j are specified completely, L is concave in $(\pi_1, \ldots, \pi_{m-1})$, and $(\hat{\pi}_1, \ldots, \hat{\pi}_m)$ is unique and can be calculated by Newton's method (Lindsay, 1995).

3 Polynomial methods for two-component mixture models

Consider a random vector $X \in \mathbb{R}^d$ with a two-component mixture density, $f(x) = \pi f_1(x) + \bar{\pi} f_2(x)$, $x \in \mathbb{R}^d$, where $\pi \in [0, 1]$ is the component weight, $\bar{\pi} \equiv 1 - \pi$, and f_1 and f_2 are completely specified densities. Without loss of generality, we assume that $f_1(x) > 0$ on $\mathbb{R}^d$ except possibly on a set of measure zero. Given a random sample $X_1, \ldots, X_n$, define $h(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_1(x_i)$ and $c(x) = f_2(x)/f_1(x)$. Then (2.2) reduces to

$$L(\pi) = \prod_{i=1}^{n} (\pi f_1(x_i) + \bar{\pi} f_2(x_i)) = h(x_1, \ldots, x_n) \prod_{i=1}^{n} (\pi + \bar{\pi} c(x_i)),$$

and the log-likelihood function,

$$\log L(\pi) = \log h(x_1, \ldots, x_n) + \sum_{i=1}^{n} \log(\pi + \bar{\pi} c(x_i)), \tag{3.1}$$

is concave in π. Define the odds ratio $u := \pi/(1-\pi)$; then $u \geq 0$ and $0 < \pi < 1$ if and only if $0 < u < \infty$. The inverse transformation from u to π is $\pi = u/(u+1)$, hence $\bar{\pi} = 1 - \pi = 1/(u+1)$. Moreover, maximization of L with respect to π is equivalent to maximization with respect to u. Denoting $c(x_i)$ by c_i and $L(\pi)$ by $\tilde{L}(u)$, we have

$$\log \tilde{L}(u) = \log h(x_1, \ldots, x_n) + \sum_{i=1}^{n} \log(u + c_i) - n \log(u+1). \tag{3.2}$$

Setting $\partial \log \tilde{L}(u)/\partial u = 0$ and denoting a nonnegative solution by $\hat{u}$, we obtain the *score equation*,

$$\frac{n}{\hat{u}+1} = \sum_{i=1}^{n} \frac{1}{\hat{u}+c_i}; \tag{3.3}$$

equivalently,

$$\left(\frac{1}{n} \sum_{i=1}^{n} (\hat{u}+c_i)^{-1} \right)^{-1} = \hat{u}+1.$$

In short, $\hat{u}+1$ equals the harmonic mean of $\{\hat{u}+c_1, \ldots, \hat{u}+c_n\}$. Applying the well-known inequality between the geometric and harmonic means, we obtain

$$\prod_{i=1}^{n} (\hat{u}+c_i)^{1/n} \geq \left(\frac{1}{n} \sum_{i=1}^{n} (\hat{u}+c_i)^{-1} \right)^{-1} = \hat{u}+1; \tag{3.4}$$

hence

$$\prod_{i=1}^{n} \left(\frac{\hat{u}+c_i}{\hat{u}+1} \right) \geq 1.$$

Denoting $\max_{1 \leq i \leq n} c_i$ by $c_{(n)}$, we find that the latter inequality implies

$$\frac{\hat{u}+c_{(n)}}{\hat{u}+1} \equiv \max_{1 \leq i \leq n} \frac{\hat{u}+c_i}{\hat{u}+1} \geq 1,$$

therefore $c_{(n)} \geq 1$. We summarize the preceding remarks as follows.

Proposition 3.1 *If $\hat{u} < \infty$ then $c_{(n)} \geq 1$; equivalently, if $c_{(n)} < 1$ then $\hat{\pi} = 1$. Moreover, if $\hat{u} < \infty$ then $c_{(1)} \leq 1$; equivalently, if $c_{(1)} > 1$ then $\hat{\pi} = 0$.*

To prove the second part of this result, we apply the above argument to the inverse odds ratio $v := (1-\pi)/\pi$, deducing that

$$\frac{n}{\hat{v}+1} = \sum_{i=1}^{n} \frac{1}{\hat{v}+{c_i}^{-1}}.$$

Proceeding as before, we find that if $\hat{v} < \infty$ then $c_{(1)} \leq 1$.

In many situations, the c_i are distinct, almost surely. This holds if the classical component densities f_1 and f_2 belong to many of the familiar continuous distributions, e.g., the classical exponential families including the multivariate normal or gamma distributions. In this situation, the inequality between the geometric and harmonic means in (3.4) is strict, and the conclusion is that $c_{(1)} < 1 < c_{(n)}$, almost surely.

We now consider in greater detail the case in which $0 < \hat{u} < \infty$. We retain the notation $c_i \equiv f_2(x_i)/f_1(x_i)$, and let

$$e_l(c_1, \ldots, c_n) = \sum_{1 \leq i_1 < \cdots < i_l \leq n} c_{i_1} \cdots c_{i_n}$$

denote the lth elementary symmetric function of $c_1, \ldots, c_n$.

Proposition 3.2 *Suppose that $0 < \hat{u} < \infty$. Then the score equation (3.3) reduces to a polynomial of degree $n-1$,*

$$\sum_{l=0}^{n-1} a_l u^l = 0 \tag{3.5}$$

where, for $l = 0, \ldots, n-1$,

$$a_l = (n-l)e_{n-l}(c_1, \ldots, c_n) - (l+1)e_{n-1-l}(c_1, \ldots, c_n). \tag{3.6}$$

Proof. By (3.3), $\hat{u}$ satisfies the equation

$$\frac{n}{u+1} = \sum_{i=1}^{n} \frac{1}{u+c_i} \equiv \frac{e_{n-1}(u+c_1, \ldots, u+c_n)}{\prod_{i=1}^{n}(u+c_i)}. \tag{3.7}$$

Therefore

$$n \prod_{i=1}^{n}(u+c_i) = (u+1)e_{n-1}(u+c_1, \ldots, u+c_n). \tag{3.8}$$

Note that

$$\prod_{i=1}^{n}(u+c_i) = \sum_{j=0}^{n} e_j(c_1, \ldots, c_n)u^{n-j}.$$

Also,

$$
\begin{aligned}
e_{n-1}(u+c_1,\ldots,u+c_n) &= \sum_{j=1}^{n}\prod_{\substack{i=1\\ i\neq j}}^{n}(u+c_i)\\
&= \sum_{j=1}^{n}\sum_{l=0}^{n-1} e_l(c_1,\ldots,c_{j-1},c_{j+1},\ldots,c_n)u^{n-1-l}\\
&= \sum_{l=0}^{n} f_l u^{n-1-l}
\end{aligned}
$$

where $f_n \equiv 0$ and

$$
f_l = \sum_{j=1}^{n} e_l(c_1,\ldots,c_{j-1},c_{j+1},\ldots,c_n), \tag{3.9}
$$

$l = 0,\ldots,n-1$. Therefore, it follows from (3.8) that

$$
n\sum_{j=0}^{n} e_j(c_1,\ldots,c_n)u^{n-j} = (u+1)\sum_{l=0}^{n-1} f_l u^{n-1-l}. \tag{3.10}
$$

By (3.9), f_l is a symmetric polynomial in $c_1,\ldots,c_n$. Moreover, f_l is homogeneous of degree l, and all monomials of degree l in $c_1,\ldots,c_n$ appear equally often in the monomial expansion of f_l; so f_l is proportional to e_l,

$$
f_l = \alpha_l\, e_l(c_1,\ldots,c_n). \tag{3.11}
$$

Evaluating both sides of (3.11) at $c_1 = \cdots = c_n = 1$, we obtain

$$
n\binom{n-1}{l} = f_l\Big|_{c_1=\cdots=c_n=1} = \alpha_l\, e_l(c_1,\ldots,c_l)\Big|_{c_1=\cdots=c_n=1} = \binom{n}{l}\alpha_l,
$$

proving that $\alpha_l = n-l$, $0 \le l \le n$.

In (3.10), the coefficient of u^n is the same on both sides of the equation. Therefore (3.10) reduces to a polynomial equation of degree $n-1$,

$$
\sum_{j=1}^{n} ne_j(c_1,\ldots,c_n)u^{n-j} = \sum_{l=0}^{n-1} f_l u^{n-1-l} + \sum_{l=1}^{n-1} f_l u^{n-l},
$$

and the latter equation reduces to

$$
\sum_{l=0}^{n-1}\left[ne_{l+1}(c_1,\ldots,c_n) - f_l - f_{l+1}\right]u^{n-1-l} = 0.
$$

Applying (3.11) with $\alpha_l = n-l$, it is straightforward to verify that

$$
ne_{l+1}(c_1,\ldots,c_n) - f_l - f_{l+1} = (l+1)e_{l+1}(c_1,\ldots,c_n) - (n-l)e_l(c_1,\ldots,c_n).
$$

The proof of (3.5) now is complete. □

By the Fundamental Theorem of Algebra, the score equation (3.5) has at most $n-1$ roots, including multiplicities. For $n \leq 5$, the resulting polynomial is of degree at most four, and then we may apply the classical theory of polynomial equations to obtain explicit formulas for the roots, thereby deriving explicit expressions for $\hat{u}$ and $\hat{\pi}$ in terms of radicals.

For general n, we can again apply the theory of polynomial equations to deduce many properties of the roots of (3.5); see Buot (2003). By applying Descartes' Rule of Signs (cf., Sturmfels, 2002), we deduce that the number of positive roots of (3.5) is bounded above by the number of sign alternations of the sequence of coefficients in (3.5). We can also obtain bounds on the number of positive roots of (3.5) which belong to any given sub-interval of the positive real line. For instance, by applying the Fourier-Budan theorem (cf., Karlin, 1968, p. 316), we have the following result: Denote by $P(u)$ the score polynomial in (3.5) and, for $t \in \mathbb{R}$, let $V(t)$ be the number of sign changes of the sequence $P(t), P'(t), P''(t), \ldots, P^{(n-1)}(t)$. For α, β with $\alpha < \beta < \infty$, the number of roots of (3.5) in the interval (α, β) is bounded above by $V(\alpha) - V(\beta)$.

In practice, numerical root calculation for (3.5) and applications of Descartes' Rule or Sturm's theorem can be carried out using the symbolic computation programs `maple` or `matlab`, as demonstrated by Sturmfels (2002), Chapter 1. Once all real roots have been computed, it is straightforward to transform these roots to the corresponding values of π. Finally, we calculate the likelihood function at each π and hence determine $\hat{\pi}$. For further details on the application of polynomial theory, and examples in which these results are applied to develop iterative estimation methods for mixture models and other statistical problems, see Buot (2003).

Although much is known about numerical methods for calculating the roots of (3.5), the statistical literature does not appear to contain any results about the more difficult problem of determining *explicit* formulas for the roots of (3.5). As mentioned above, for sample sizes of five or fewer, explicit formulas for $\hat{\pi}$ can be obtained in terms of radicals. For larger sample sizes, results of Sturmfels (2000) and earlier authors imply explicit infinite series representations for the solutions of (3.5) in terms of the coefficients a_l. Consider, for example, the case of a two-component mixture model with a random sample of size six. Then (3.5) is a fifth-degree polynomial equation. Explicit series expansions for its roots are of the following type (cf. Sturmfels, 2000, Section 1.5),

$$u = \sum_{i,j,k,l \geq 0} \frac{(-1)^{1+2i+3j+4k+5l}\,(2i+3j+4k+5l)!}{i!\,j!\,k!\,l!\,(1+i+2j+3k+4l)!} \frac{a_0^{1+i+2j+3k+4l} a_2^i a_3^j a_4^k a_5^l}{a_1^{1+2i+3j+4k+5l}},$$

subject to the a_j being such that the series converges. These formulas provide explicit, non-iterative, solutions to the score equation, hence explicit formulas for the maximum likelihood estimator $\hat{\pi}$, and therefore suggest a way toward analytic treatments of the exact, small sample, distribution of $\hat{\pi}$.

4 Polynomial methods for general mixture models

Consider the general mixture model with density function (2.1), where $m \geq 3$. We denote by $\boldsymbol{\pi}$ the vector $(\pi_1, \dots, \pi_m)$ of component proportions. Generalizing Proposition 3.1, we can derive results pertaining to the case in which $\hat{\boldsymbol{\pi}}$ falls on the boundary of the simplex $\mathcal{S}_m$; the details of these results are similar to the foregoing, and are left to the reader. Therefore, we restrict our attention to the case in which $\hat{\boldsymbol{\pi}}$ falls within the interior of $\mathcal{S}_m$.

Without loss of generality, we assume that $f_1(x) > 0$ on $\mathbb{R}^d$ except possibly on a set of measure zero. Given a random sample $X_1, \dots, X_n$, we write the likelihood function (2.2) in the form

$$L(\boldsymbol{\pi}) = \prod_{i=1}^{n} \Big[\sum_{j=1}^{m} \pi_j f_j(x_i)\Big] = \prod_{i=1}^{n} f_1(x_i) \Big[\sum_{j=1}^{m} \pi_j \frac{f_j(x_i)}{f_1(x_i)}\Big]. \tag{4.1}$$

Using the notation $\boldsymbol{x} = (x_1, \dots, x_n)$, define $h(\boldsymbol{x}) = \prod_{i=1}^{n} f_1(x_i)$ and

$$c_{i,j} = f_j(x_i)/f_1(x_i), \tag{4.2}$$

$j = 1, \dots, m$, $i = 1, \dots, n$. Then (4.1) becomes

$$L(\boldsymbol{\pi}) = h(\boldsymbol{x}) \prod_{i=1}^{n} \Big[\sum_{j=1}^{m} c_{i,j}\pi_j\Big]. \tag{4.3}$$

We introduce the change of variables from $\boldsymbol{\pi}$ to $\boldsymbol{u} = (u_1, \dots, u_m)$ where

$$\pi_j = u_j/(u_1 + \cdots + u_m), \tag{4.4}$$

$j = 1, \dots, m$. This is a one-to-one transformation from the interior of $\mathcal{S}_m$ to the positive orthant $\mathbb{R}_+^{m-1}$, with inverse transformation $u_j = \pi_j/\pi_m$, $j = 1, \dots, m$; in particular, $u_m \equiv 1$.

For $i = 1, \dots, n$, define

$$P_i(\boldsymbol{u}) = \sum_{j=1}^{m} c_{i,j} u_j, \tag{4.5}$$

Applying (4.4) to (4.3) and denoting the resulting function by $\tilde{L}(\boldsymbol{u})$, we obtain

$$\log \tilde{L}(\boldsymbol{u}) = \log h(\boldsymbol{x}) + \sum_{i=1}^{n} \log P_i(\boldsymbol{u}) - n \log \Big(\sum_{j=1}^{m} u_j\Big). \tag{4.6}$$

Since the transformation (4.4) is one-to-one and the logarithm function is strictly increasing, maximization of (4.1) over $\mathcal{S}_m$ is equivalent to maximization of (4.6) over $\mathbb{R}_+^{m-1}$. Differentiating $\log \tilde{L}$ with respect to each u_k and setting all derivatives equal to zero, we obtain the system of equations

$$\sum_{i=1}^{n} \frac{c_{i,k}}{P_i(\boldsymbol{u})} - \frac{n}{\sum_{j=1}^{m} u_j} = 0, \tag{4.7}$$

$k = 1, \ldots, m-1$. By clearing denominators, we rewrite this system in polynomial form: For $k = 1, \ldots, m-1$,

$$n \prod_{i=1}^{n} P_i(\boldsymbol{u}) - \Big(\sum_{j=1}^{m} u_j\Big)\Big(\sum_{i=1}^{n} c_{i,k} \prod_{\substack{l=1 \\ l \neq i}}^{n} P_l(\boldsymbol{u})\Big) = 0. \tag{4.8}$$

Let $\mathbb{C}^*$ denote the set of non-zero complex numbers. Call a solution $\boldsymbol{u}$ of (4.8) *isolated* if there is a neighborhood of $\boldsymbol{u}$ containing no other solutions. We wish to determine the number of isolated solutions of (4.8) in $(\mathbb{C}^*)^{m-1}$, i.e., all solutions for which $u_j \neq 0$, $j = 1, \ldots, m-1$; this explains why we treated the boundary cases earlier. We call the system (4.8) *generic* if all its *facial resultants* are non-zero; here, "facial resultants" are certain determinants defined in terms of the coefficients of the system (Huber and Sturmfels, 1995). We note also that by applying the results of Sturmfels (1994), we can find explicit characterizations of the resultants for the system (4.8); however, such results will not be needed here since it suffices to know only that the resultants are nontrivial polynomials in the coefficients of (4.8).

By applying a fundamental theorem of Bernstein (1975) (cf. Cox, et al., 1998; Sturmfels, 1998) we shall obtain our main result, stated as follows.

Theorem 4.1 *Suppose that the system (4.8) has a finite number of isolated solutions in* $(\mathbb{C}^*)^{m-1}$*. Then there exist at most* n^{m-1} *such solutions. Moreover, if (4.8) is generic then, almost surely, it has exactly* n^{m-1} *solutions in* $(\mathbb{C}^*)^{m-1}$*.*

Examples of component densities for which (4.8) is generic (almost surely) are the multivariate normal and gamma distributions, or any continuous member of the exponential family of distributions. For these distributions, the coefficients of (4.8), being polynomials in the $c_{i,j}$ of (4.2), are random variables continuous on their support sets, and the closure of their support sets is the full sample space $\mathbb{R}^d$. The facial resultants, being nontrivial polynomials in the $c_{i,j}$, then are non-zero, almost surely. On the other hand, if the component densities are discrete then in such cases the data $x_1, \ldots, x_n$ may generate coefficients $c_{i,j}$ which satisfy polynomial equations with positive probability, and then some facial resultants may be identically zero; in such cases, it is possible to have fewer than n^{m-1} solutions.

For $m = 3$, Theorem 4.1 and the remarks *infra* are due to Buot (2003).

Returning to (4.8), we apply a change of indices to obtain

$$\prod_{i=1}^{n} P_i(\boldsymbol{u}) \equiv \prod_{i=1}^{n} \Big(\sum_{j_i=1}^{m} c_{i,j_i} u_{j_i} \Big) = \sum_{j_1,\ldots,j_n=1}^{m} \prod_{l=1}^{n} c_{l,j_l} u_{j_l}.$$

Rearranging terms in this sum so as to collect powers of $u_1, u_2, \ldots$, we find that this last expression equals

$$\sum_{r_1+\cdots+r_m=n} \Bigg[\sum_{(j_1,\ldots,j_n)} \prod_{l=1}^{n} c_{l,j_l} \Bigg] u_1^{r_1} u_2^{r_2} \cdots u_m^{r_m}, \tag{4.9}$$

where the outer sum is over all nonnegative integers $r_1, \ldots, r_m$ such that $r_1 + \cdots + r_m = n$ and the inner sum is over all positive integers $j_1, \ldots, j_n$ such that the set $\{j_1, \ldots, j_n\}$ consists of r_1 ones, r_2 twos, ..., r_m m's. Introduce the multi-index notation $\boldsymbol{r} = (r_1, \ldots, r_m)$, $|\boldsymbol{r}| = r_1 + \cdots + r_m$, and $\boldsymbol{u}^{\boldsymbol{r}} = u_1^{r_1} u_2^{r_2} \cdots u_m^{r_m}$. Let $C = (c_{i,j})$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, be the $n \times m$ matrix with (i,j)th entry $c_{i,j}$, and denote by $\gamma_{\boldsymbol{r}}(C)$ the term in square brackets in (4.9); then (4.9) reduces to

$$\prod_{i=1}^{n} P_i(\boldsymbol{u}) = \sum_{|\boldsymbol{r}|=n} \gamma_{\boldsymbol{r}}(C) \boldsymbol{u}^{\boldsymbol{r}}. \tag{4.10}$$

For $i = 1, \ldots, n$, let $C_i = (c_{l,j})$, $l = 1, \ldots, i-1, i+1, \ldots, n$, $j = 1, \ldots, m$, be the $(n-1) \times m$ matrix obtained by deleting from C the ith row. For nonnegative integers $s_1, \ldots, s_m$, let $\boldsymbol{s} = (s_1, \ldots, s_m)$ and define

$$\gamma_{\boldsymbol{s}}(C_i) = \sum_{(j_1,\ldots,j_{i-1},j_{i+1},\ldots,j_{n-1})} \prod_{\substack{l=1 \\ l \neq i}}^{n} c_{l,j_l}, \tag{4.11}$$

where the sum is over all positive integers $j_1, \ldots, j_{i-1}, j_{i+1}, \ldots, j_n$ such that the set $\{j_1, \ldots, j_{i-1}, j_{i+1}, \ldots, j_n\}$ consists of s_1 ones, s_2 twos, ..., s_m m's. Similar to (4.10), we then obtain

$$\prod_{\substack{l=1 \\ l \neq i}}^{n} P_l(\boldsymbol{u}) = \sum_{|\boldsymbol{s}|=n-1} \gamma_{\boldsymbol{s}}(C_i) \boldsymbol{u}^{\boldsymbol{s}}. \tag{4.12}$$

Collecting together (4.8) - (4.12), we find that for $k = 1, \ldots, m-1$,

$$n \sum_{|\boldsymbol{r}|=n} \gamma_{\boldsymbol{r}}(C) \boldsymbol{u}^{\boldsymbol{r}} - \Big(\sum_{j=1}^{m} u_j \Big) \sum_{i=1}^{n} c_{i,k} \sum_{|\boldsymbol{s}|=n-1} \gamma_{\boldsymbol{s}}(C_i) \boldsymbol{u}^{\boldsymbol{s}} = 0. \tag{4.13}$$

In the sequel, we denote by $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{m-1}$ the standard basis for $\mathbb{R}^{m-1}$, i.e., $\boldsymbol{e}_k = (0, \ldots, 0, 1, 0, \ldots, 0)$, the vector with zero coordinates in all entries except for a 1 in the kth coordinate; and we define $\boldsymbol{e}_m = (0, \ldots, 0)$, the zero vector in $\mathbb{R}^{m-1}$. With the usual notation $\delta_{j,k}$ for Kronecker's delta function, we also denote by $(\boldsymbol{e}_j, \delta_{j,m})$ the m-dimensional vector whose first $m-1$ coordinates are the same as $\boldsymbol{e}_j$, and whose last coordinate is $\delta_{j,m}$. With this notation, $u_j \equiv \boldsymbol{u}^{(\boldsymbol{e}_j, \delta_{j,m})}$, $j = 1, \ldots, m$, so the kth polynomial in (4.13) is

$$n \sum_{|\boldsymbol{r}|=n} \gamma_{\boldsymbol{r}}(C) \boldsymbol{u}^{\boldsymbol{r}} - \Big(\sum_{j=1}^{m} \boldsymbol{u}^{(\boldsymbol{e}_j, \delta_{j,m})} \Big) \Big(\sum_{i=1}^{n} c_{i,k} \sum_{|\boldsymbol{s}|=n-1} \gamma_{\boldsymbol{s}}(C_i) \boldsymbol{u}^{\boldsymbol{s}} \Big). \tag{4.14}$$

Multiplying the sums and collecting monomials in $\boldsymbol{u}$, we find that the coefficient of $\boldsymbol{u}^{\boldsymbol{r}}$ in (4.14) is

$$n\gamma_{\boldsymbol{r}}(C) - \sum_{j=1}^{m} \sum_{\substack{|\boldsymbol{s}|=n-1 \\ \boldsymbol{s}+(\boldsymbol{e}_j, \delta_{j,m})=\boldsymbol{r}}} \sum_{i=1}^{n} c_{i,k} \gamma_{\boldsymbol{s}}(C_i). \tag{4.15}$$

To calculate the number of simultaneous solutions to the system of score equations (4.13), we need to determine Θ_k, the *carrier set* of the kth equation in the system, i.e., the set of exponents $\boldsymbol{r}$ for which (4.15) is non-zero. Denoting by $\mathbb{N}_0$ the set of nonnegative integers, we have the following result.

Proposition 4.2 *For $k = 1, \ldots, m-1$, the carrier set of the system (4.13) is the set of lattice points*

$$\Theta_k = \{(i_1, \ldots, i_{m-1}) \in \mathbb{N}_0^{m-1} : i_1 + \cdots + i_{m-1} \le n\} \setminus \{n\boldsymbol{e}_k\}. \tag{4.16}$$

Proof. By (4.10),

$$\sum_{|\boldsymbol{r}|=n} \gamma_{\boldsymbol{r}}(C) \boldsymbol{u}^{\boldsymbol{r}} = \prod_{l=1}^{n} P_l(\boldsymbol{u}) = P_i(\boldsymbol{u}) \prod_{\substack{l=1 \\ l\neq i}}^{n} P_i(\boldsymbol{u}),$$

$i = 1, \ldots, n$; therefore

$$n \sum_{|\boldsymbol{r}|=n} \gamma_{\boldsymbol{r}}(C) \boldsymbol{u}^{\boldsymbol{r}} = \sum_{i=1}^{n} P_i(\boldsymbol{u}) \prod_{\substack{l=1 \\ l\neq i}}^{n} P_i(\boldsymbol{u}).$$

It now follows from (4.5) and (4.12) that

$$\begin{aligned} n \sum_{|\boldsymbol{r}|=n} \gamma_{\boldsymbol{r}}(C) \boldsymbol{u}^{\boldsymbol{r}} &= \sum_{i=1}^{n} \Big[\sum_{j=1}^{m} c_{i,j} \boldsymbol{u}^{(\boldsymbol{e}_j, \delta_{j,m})} \Big] \Big[\sum_{|\boldsymbol{s}|=n-1} \gamma_{\boldsymbol{s}}(C_i) \boldsymbol{u}^{\boldsymbol{s}} \Big] \\ &= \sum_{|\boldsymbol{r}|=n} \Big[\sum_{j=1}^{m} \sum_{\substack{|\boldsymbol{s}|=n-1 \\ \boldsymbol{s}+(\boldsymbol{e}_j, \delta_{j,m})=\boldsymbol{r}}} \sum_{i=1}^{n} c_{i,j} \gamma_{\boldsymbol{s}}(C_i) \Big] \boldsymbol{u}^{\boldsymbol{r}}. \end{aligned} \tag{4.17}$$

Comparing the coefficient of $\boldsymbol{u}^{\boldsymbol{r}}$ on both sides of (4.17), we see that

$$n\gamma_{\boldsymbol{r}}(C) = \sum_{\substack{j=1 \\ \boldsymbol{s}+(\boldsymbol{e}_j,\delta_{j,m})=\boldsymbol{r}}}^{m} \sum_{|\boldsymbol{s}|=n-1} \sum_{i=1}^{n} c_{i,j}\gamma_{\boldsymbol{s}}(C_i)$$

By (4.15), the coefficient of $\boldsymbol{u}^{\boldsymbol{r}}$ in the kth score polynomial is

$$\sum_{\substack{j=1 \\ \boldsymbol{s}+(\boldsymbol{e}_j,\delta_{j,m})=\boldsymbol{r}}}^{m} \sum_{|\boldsymbol{s}|=n-1} \sum_{i=1}^{n} c_{i,j}\gamma_{\boldsymbol{s}}(C_i) - \sum_{\substack{j=1 \\ \boldsymbol{s}+(\boldsymbol{e}_j,\delta_{j,m})=\boldsymbol{r}}}^{m} \sum_{|\boldsymbol{s}|=n-1} \sum_{i=1}^{n} c_{i,k}\gamma_{\boldsymbol{s}}(C_i)$$
$$\equiv \sum_{\substack{j=1,j\neq k \\ \boldsymbol{s}+(\boldsymbol{e}_j,\delta_{j,m})=\boldsymbol{r}}}^{m} \sum_{|\boldsymbol{s}|=n-1} \sum_{i=1}^{n} (c_{i,j} - c_{i,k})\gamma_{\boldsymbol{s}}(C_i). \quad (4.18)$$

To determine the coefficient of u_k^n, set $r_k = n$ and $r_j = 0$ otherwise. Because all coordinates of the vectors $\boldsymbol{s}$, $\boldsymbol{r}$, and $(\boldsymbol{e}_j, \delta_{j,m})$ are nonnegative integers, the constraint $\boldsymbol{s} + (\boldsymbol{e}_j, \delta_{j,m}) = \boldsymbol{r}$ implies that $s_j \leq r_j$, $j = 1, \ldots, m$. Since $|\boldsymbol{s}| = n - 1$, it follows that $s_k = n - 1$ and $s_j = 0$ otherwise. Therefore the set of "admissible" j in (4.18) satisfies $(\boldsymbol{e}_j, \delta_{j,m}) = (\boldsymbol{e}_k, 0)$, which is valid if and only if $j = k$. Since the term $j = k$ does not appear in (4.18), the set of admissible j is empty. Therefore, the coefficient of u_k^n is zero.

Next, consider the monomials $\boldsymbol{u}^{\boldsymbol{r}}$ where $\boldsymbol{r} = (r_1, \ldots, r_m)$ with $r_k < n$. By (4.18), the set of admissible j is non-empty, for it contains at least one element of the set $\{1, \ldots, k-1, k+1, \ldots, m\}$; therefore (4.18) contains at least one summand, and we shall show that (4.18) is a nontrivial polynomial in the $c_{i,j}$. To that end, fix the terms $c_{i,k}$ and consider (4.18) as a polynomial function of the $c_{i,j}$, $j \neq k$; replacing each of these $c_{i,j}$ by $c_{i,j} + c_{i,k}$, the right-hand side of (4.18) becomes a sum of terms $c_{i,j}\gamma_{\boldsymbol{s}}(\tilde{C}_i)$, where $\tilde{C}_i = (c_{l,j} + c_{l,k})$ is the matrix obtained by replacing each entry $c_{l,j}$ in C_i by $c_{l,j} + c_{l,k}$, and $\gamma_{\boldsymbol{s}}(\tilde{C}_i)$ is defined as in (4.11). However, each term $c_{i,j}\gamma_{\boldsymbol{s}}(\tilde{C}_i)$, clearly, is a sum of monomials in the $c_{i,j}$, hence is a nontrivial polynomial. Therefore (4.18) is also a nontrivial polynomial in the $c_{i,j}$, and the proof is complete. □

From (4.16), it follows that the carrier set Θ_k is convex. By counting the number of nonnegative integer solutions to the inequality $i_1 + \ldots + i_{m-1} \leq n$, we also find that Θ_k has cardinality $\ell = \binom{n+m-1}{m-1} - 1$. Let $\boldsymbol{v}_{1,k}, \ldots, \boldsymbol{v}_{\ell,k}$ be an enumeration of all vectors in Θ_k, and construct the convex hull

$$\mathcal{C}_k = \{\lambda_1\boldsymbol{v}_{1,k} + \cdots + \lambda_\ell\boldsymbol{v}_{\ell,k} : \lambda_i \geq 0, \lambda_1 + \cdots + \lambda_\ell = 1\}.$$

In the terminology of algebraic geometry (Cox, et al., 1998, Chapter 7; Sturmfels, 1998-2002), $\mathcal{C}_k$ is the *Newton polytope* of the polynomial (4.14).

By (4.16), we have $\mathcal{C}_k = \mathcal{P} \setminus \mathcal{P}_k$ where

$$\mathcal{P} = \{(w_1, \ldots, w_{m-1}) \in \mathbb{R}^{m-1} : w_j \geq 0, 1 \leq j \leq m-1, \sum_{j=1}^{m-1} w_j \leq n\}, \quad (4.19)$$

and

$$\mathcal{P}_k = \{(w_1, \ldots, w_{m-1}) \in \mathcal{P} : n-1 \leq w_k \leq n\} \quad (4.20)$$

is the unit subpolyhedron of $\mathcal{P}$ located at the vertex $n\boldsymbol{e}_k$. In short, $\mathcal{C}_k$ is the polyhedron obtained by excising from $\mathcal{P}$ a subpolyhedron located at an extreme point. Therefore $\mathrm{Vol}(\mathcal{C}_k)$, the volume of $\mathcal{C}_k$, equals $\mathrm{Vol}(\mathcal{P}) - \mathrm{Vol}(\mathcal{P}_k)$.

Given the Newton polytopes $\mathcal{C}_{i_1}, \ldots, \mathcal{C}_{i_k}$, we construct their *Minkowski sum*, $\mathcal{C}_{i_1} + \cdots + \mathcal{C}_{i_k} := \{\boldsymbol{v}_1 + \cdots + \boldsymbol{v}_k : \boldsymbol{v}_j \in \mathcal{C}_{i_j}, j = 1, \ldots, k\}$. Then, we need to calculate the *mixed volume* of the polytopes $\mathcal{C}_1, \ldots, \mathcal{C}_{m-1}$,

$$\mathcal{M}(\mathcal{C}_1, \ldots, \mathcal{C}_{m-1}) := \sum_{k=1}^{m-1} (-1)^{m-1-k} \sum_{1 \leq i_1 < \cdots < i_k \leq m-1} \mathrm{Vol}(\mathcal{C}_{i_1} + \cdots + \mathcal{C}_{i_k}). \quad (4.21)$$

Proposition 4.3 *For $m \geq 3$, $\mathcal{M}(\mathcal{C}_1, \ldots, \mathcal{C}_{m-1}) = n^{m-1}$.*

Proof. Fix $i_1, \ldots, i_k$ where $1 \leq i_1 < \cdots < i_k \leq m-1$. By geometrical considerations as at (4.19) and (4.20), we find that $\mathcal{C}_{i_1} + \cdots + \mathcal{C}_{i_k}$ is a large polyhedron from which unit subpolyhedrons corresponding to the vertices $n\boldsymbol{e}_{i_1}, \ldots, n\boldsymbol{e}_{i_k}$ have been excised:

$$\mathcal{C}_{i_1} + \cdots + \mathcal{C}_{i_k} = k\mathcal{P} \setminus \bigcup_{j=1}^{k} E_{i_k} \quad (4.22)$$

where $k\mathcal{P} = \{k\boldsymbol{v} : \boldsymbol{v} \in \mathcal{P}\}$ and $E_{i_j} = \{(w_1, \ldots, w_{m-1}) \in k\mathcal{P} : nk - 1 \leq w_{i_j} \leq nk\}$. It is not difficult to see that E_{i_j} has the same volume as E_j; therefore $\mathrm{Vol}(\mathcal{C}_{i_1} + \cdots + \mathcal{C}_{i_k})$ does not depend on $i_1, \ldots, i_k$. It follows that $\mathcal{C}_{i_1} + \cdots + \mathcal{C}_{i_k}$ and $\mathcal{C}_1 + \cdots + \mathcal{C}_k$ have the same volume and then, by (4.3), we have

$$\mathcal{M}(\mathcal{C}_1, \ldots, \mathcal{C}_{m-1}) = \sum_{k=1}^{m-1} (-1)^{m-1-k} \binom{m-1}{k} \mathrm{Vol}(\mathcal{C}_1 + \cdots + \mathcal{C}_k).$$

Applying (4.22) to the polytopes $\mathcal{C}_1, \ldots, \mathcal{C}_k$ and noting that each subpolyhedron E_j has the same volume, we obtain

$$\mathrm{Vol}(\mathcal{C}_1 + \cdots + \mathcal{C}_k) = \mathrm{Vol}(k\mathcal{P}) - k\mathrm{Vol}(E_1) = k^{m-1}\mathrm{Vol}(\mathcal{P}) - k\mathrm{Vol}(E_1).$$

Simple geometrical considerations, or a change of variables in the coordinates defining E_1, show that E_1 has the same volume as the unit polyhedron $E_0 =$

$\{(w_1, \ldots, w_{m-1}) \in \mathbb{R}^{m-1} : w_j \geq 0, 1 \leq j \leq m-1, \sum_{j=1}^{m-1} w_j \leq 1\}$. Therefore,

$$\text{Vol}(E_1) = \int \cdots \int_{w_j \geq 0,\ \sum_{j=1}^{m-1} w_j \leq 1} dw_1 \cdots dw_{m-1} = \frac{1}{(m-1)!},$$

where the last equality follows from results basic to the Dirichlet distributions. Similarly,

$$\text{Vol}(\mathcal{P}) = \int \cdots \int_{w_j \geq 0,\ \sum_{j=1}^{m-1} w_j \leq n} dw_1 \cdots dw_{m-1} = \frac{n^{m-1}}{(m-1)!}.$$

Collecting together these results, we have

$$\mathcal{M}(\mathcal{C}_1, \ldots, \mathcal{C}_{m-1}) = \frac{1}{(m-1)!} \sum_{k=1}^{m-1} (-1)^{m-1-k} \binom{m-1}{k} (n^{m-1} k^{m-1} - k). \tag{4.23}$$

Recall the combinatorial identity

$$\sum_{k=1}^{m-1} (-1)^{m-1-k} \binom{m-1}{k} k \equiv 0,$$

proved by differentiating the binomial expansion of $(1+t)^{m-1}$ with respect to t and evaluating the outcome at $t = -1$. Also, by elementary properties of the Stirling numbers of the second kind (cf., Stanley, 1986, pp. 33–34),

$$\sum_{k=1}^{m-1} (-1)^{m-1-k} \binom{m-1}{k} k^{m-1} \equiv \sum_{k=0}^{m-1} (-1)^{m-1-k} \binom{m-1}{k} k^{m-1} = (m-1)!.$$

Applying these identities to (4.23), we obtain the desired conclusion. □

We have now assembled the ingredients necessary to complete the proof of Theorem 4.1. According to Bernstein's theorem, the number of isolated solutions of (4.8) in $(\mathbb{C}^*)^{m-1}$ is bounded above by the mixed volume $\mathcal{M}(\mathcal{C}_1, \ldots, \mathcal{C}_{m-1})$, and the upper bound is exact for generic systems. Therefore, by Proposition 4.21, we obtain the upper bound of n^{m-1} for any mixture model. Further, for cases in which the system (4.8) is generic, that upper bound is, almost surely, the exact number of solutions.

5 Homotopy continuation for a three-component mixture model

We have noted before that there exists at most one maximum likelihood estimate $\hat{\boldsymbol{\pi}}$ in $\mathcal{S}_m$, and the solution may be calculated using Newton's method.

Equivalently, the system (4.8) has at most one solution in $\mathbb{R}_+^{m-1}$. Thus, in the case of the mixture models considered here, homotopy continuation methods are neither superior nor inferior to classical iterative methods (by contrast, for other statistical problems, we shall demonstrate in future articles that the homotopy continuation approach provides significant advantages over some well-known classical iterative methods). Still, it is instructive to consider an example in which homotopy methods are applied to estimating the components in a mixture model.

Denote by $N(\mu, \sigma^2)$ the normal distribution on $\mathbb{R}$ with mean μ and variance σ^2. Let f_1, f_2, and f_3 be the density functions of the normal distributions $N(-1,1)$, $N(0,1)$, and $N(1,1)$, respectively. Also, let X be a random variable with a three-component mixture density $f(x) = \pi_1 f_1(x) + \pi_2 f_2(x) + \pi_3 f_3(x)$, $x \in \mathbb{R}$. Note that $f_1(x) > 0$, $x \in \mathbb{R}$. Also, the densities f_1, f_2, and f_3 belong to the exponential family of continuous densities, and the system of equations (4.8) is generic, almost surely, so the hypotheses of Theorem 4.1 are satisfied. Therefore, almost surely, (4.8) has exactly twenty-five solutions in $(\mathbb{C}^*)^2$.

Suppose that the vector of mixing proportions is $\boldsymbol{\pi} = (0.5, 0.3, 0.2)$. Using the statistical package R (Ihaka and Gentleman, 1996), we simulated a random sample of five observations from the mixture random variable X, obtaining the simulated data -1.64, -0.97, -0.44, 0.32, and 1.74. The system (4.8) then is obtained in the form $g_1(\boldsymbol{u}) = 0$, $g_2(\boldsymbol{u}) = 0$, $\boldsymbol{u} = (u_1, u_2)$ where

$$\begin{aligned} g_1(\boldsymbol{u}) = & -4.35 - 58.78u_1 - 127.96u_1^2 + 24.43u_1^3 + 29.95u_1^4 \\ & -42.22u_2 - 286.60u_1u_2 - 145.98u_1^2u_2 + 141.63u_1^3u_2 + 8.67u_1^4u_2 \\ & -125.98u_2^2 - 328.34u_1u_2^2 + 94.42u_1^2u_2^2 + 37.19u_1^3u_2^2 \\ & -133.88u_2^3 - 72.35u_1u_2^3 + 33.32u_1^2u_2^3 \\ & -47.48u_2^4 - 1.54u_1u_2^4 - 5.50u_2^5, \end{aligned}$$

$$\begin{aligned} g_2(\boldsymbol{u}) = & -0.36 - 5.74u_1 - 11.71u_1^2 + 9.06u_1^3 - 1.54u_1^4 - 1.92u_1^5 \\ & -3.62u_2 - 25.00u_1u_2 + 7.14u_1^2u_2 + 19.76u_1^3u_2 - 8.22u_1^4u_2 \\ & -10.13u_2^2 - 13.19u_1u_2^2 + 36.62u_1^2u_2^2 - 7.36u_1^3u_2^2 \\ & -7.68u_2^3 + 13.86u_1u_2^3 + 0.34u_1^2u_2^3 + 0.03u_2^4 + 1.22u_1u_2^4. \end{aligned}$$

Clearly, the carrier sets of g_1 and g_2 are as stated in Proposition 4.2. Solving the system $g_1(\boldsymbol{u}) = g_2(\boldsymbol{u}) = 0$ using `PHCpack`, a homotopy continuation algorithm for solving systems of polynomial equations (Verschelde, 1999), we obtain $\boldsymbol{u} = (2.086, 1.035)$. Applying the transformation (4.4), we obtain $\hat{\boldsymbol{\pi}} = (0.51, 0.25, 0.24)$, the unique maximum likelihood estimate of $\boldsymbol{\pi}$.

Because the facial resultants of the system $g_1(\boldsymbol{u}) = g_2(\boldsymbol{u}) = 0$ are non-zero, we know that there exists exactly twenty-five solutions of the system in $(\mathbb{C}^*)^2$. At first glance, it is remarkable to us that the polynomials g_1 and g_2 have no more or no fewer solutions. Even more remarkable is that exactly one of these twenty-five solutions lies in the orthant $\mathbb{R}^2_+$. In addition to finding the unique solution $\hat{\boldsymbol{u}} \in \mathbb{R}^2_+$, the package `PHCpack` also locates all solutions in $(\mathbb{C}^*)^2$. For the system given above, `PHCpack` converged twenty-one times to real solutions; for example, a solution falling outside the orthant $\mathbb{R}^2_+$ is $\boldsymbol{u} = (7.845, -4.291)$.

In maximum likelihood estimation of $\boldsymbol{\pi}$, it is the real roots which are of interest to statisticians; for this reason, in future papers, we shall place greater emphasis on homotopy continuation methods for locating real roots to systems of polynomial equations. Nevertheless, as explained by Sturmfels (2002), counting the number of real roots of a system is more difficult than counting all complex roots; this is the case even in the classical one-dimensional case, where the Fundamental Theorem of Algebra provides a definitive answer for the number of complex solutions of a single-variable polynomial equation, whereas no real analog of the Fundamental Theorem is available. We also remark that the ability of `PHCpack` to locate complex solutions raises the issue of meaningful statistical interpretation for the complex roots of a system of maximum likelihood equations. Despite an exhaustive search of the literature, it appears that such considerations have not been carried out.

Acknowledgments. This article was the basis for a talk given at the December, 2003 workshop on Computational Algebraic Statistics, held at the American Institute of Mathematics in Palo Alto, California. We are grateful to Stephen Fienberg and Bernd Sturmfels for the invitation to participate in the workshop.

References

[1] Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.*, **1**, 135–141.

[2] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* Third edition. Wiley, New York.

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

[4] Buot, M.-L. G. (2003). *Genetic Algorithms and Maximum Likelihood Estimation.* Doctoral Dissertation, University of Virginia.

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

[6] Diaconis, P. and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.*, **26**, 363–397.

[7] Drton, M. and Richardson, T. S. (2004). Multimodality of the likelihood

in the bivariate seemingly unrelated regressions model. *Biometrika*, to appear.

[8] Fienberg, S. E. and Meyer, M. M. (1983). Loglinear models and categorical data analysis with psychometric and econometric applications. *J. Econometrics*, **22**, 191–214.

[9] Hoferkamp, C. L., and Peddada, S. D. (2002). Parameter estimation in linear models with heteroscedastic variances subject to order restrictions. *J. Multivariate Anal.*, **82**, 65–87.

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

[11] Ihaka, R. and Gentleman, R. (1996). *R*: A language for data analysis and graphics. *J. Comput. Graphical Statist.*, **5**, 299–314.

[12] Karlin, S. (1968). *Total Positivity.* Stanford University Press, Stanford, CA.

[13] Lindsay, B. (1995). *Mixture Models: Theory, Geometry and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 5. Institute of Mathematical Statistics, Hayward, CA.

[14] Little, R. J. A. (2002). *Statistical Analysis with Missing Data.* Second edition. Wiley-Interscience, Hoboken, NJ.

[15] Stanley, R. (1986). *Enumerative Combinatorics.* Volume 1. Wadsworth, Belmont, CA.

[16] Sturmfels, B. (1994). On the Newton polytope of the resultant. J. Algebraic Combin., **3**, 207–236.

[17] Sturmfels, B. (1998) Polynomial equations and convex polytopes. *Amer. Math. Monthly*, **105**, 907–922.

[18] Sturmfels, B. (2000). Solving algebraic equations in terms of $\mathcal{A}$-hypergeometric Series. *Discrete Mathematics*, **210**, 171–181.

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

[20] Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions.* Wiley, New York.

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