Ryan ELMORE, Peter HALL & Amnon NEEMAN

An application of classical invariant theory to identifiability in nonparametric mixtures

AN APPLICATION OF CLASSICAL INVARIANT THEORY TO IDENTIFIABILITY IN NONPARAMETRIC MIXTURES

by Ryan ELMORE, Peter HALL & Amnon NEEMAN[(*)]

0. Introduction.

It is known that problems on the multivariate mixing model can be reduced to algebraic geometry. We begin the introduction by reminding the reader of the statistical problem. Then we restate it as a problem about some map of algebraic varieties. We then proceed to state our main theorem. Only after that, at the end of the introduction, do we refer the reader to the literature. We discuss briefly the way the statistical question has usually been reformulated (our way of translating it into algebraic geometry is very slightly different from the standard), and then we compare our result to previous work.

To illustrate the statistical problem we will look at a simple example. The example we choose is parametric. The statistician reading the article will object that the example misses the whole point of what is new and exciting about our result. The main theorem we prove is a statement that can be immediately applied both to the parametric and the non-parametric situations. In the parametric case the statistician will not be

very impressed; parametric mixture models are well understood. There is literature on them going back decades. The remarkable innovation here is that our result also applies to the non-parametric situation. Nevertheless, we want a non-technical example that will explain to the non-expert what this is all about. For this it is easier to have an example with finitely many parameters.

Suppose that, on the island of Cyprus, we take a random sample of 2 000 men in their mid-forties. For each person we measure (1) the person's height; (2) the person's systolic blood pressure; and (3) the color of the person's eyes.

The island of Cyprus has two populations on it: Greek Cypriots and Turkish Cypriots. Let us assume that the proportion of the Greek Cypriots is π_1 and the proportion of the Turkish Cypriots π_2. (Of course, $\pi_1 + \pi_2 = 1$.)

Within each population, it may not be unreasonable to assume that the height, systolic blood pressure and eye color are independent of each other. Let us suppose that the proportion of Greek (respectively Turkish) Cypriots with height < 167cm is x_1 (respectively, x_2). Assume too that the proportion of Greek (respectively Turkish) Cypriots with systolic blood pressure < 120 is y_1 (respectively, y_2). Finally, suppose that the proportion of Greek (respectively Turkish) Cypriots with blue eyes is z_1 (respectively, z_2).

With all these assumptions, a person picked at random will have probability

$$\pi_1 x_1 y_1 z_1 + \pi_2 x_2 y_2 z_2$$

of being < 167cm tall, having systolic blood pressure < 120 and blue eyes. The person would have a probability of

$$\pi_1 x_1 z_1 + \pi_2 x_2 z_2$$

of being < 167cm tall and having blue eyes; and so on. In other words, from the sample we can construct estimates of the seven numbers

$$\pi_1 x_1 y_1 z_1 + \pi_2 x_2 y_2 z_2, \quad \pi_1 x_1 y_1 + \pi_2 x_2 y_2,$$
$$\pi_1 x_1 z_1 + \pi_2 x_2 z_2, \quad \pi_1 y_1 z_1 + \pi_2 y_2 z_2$$
$$\pi_1 x_1 + \pi_2 x_2, \qquad \pi_1 y_1 + \pi_2 y_2, \qquad \pi_1 z_1 + \pi_2 z_2.$$

This model is called the multivariate mixture model. The inversion problem (or *identifiability* problem) for the model asks whether, from the seven

numbers above that we can estimate, we can solve for the seven numbers x_1, x_2, y_1, y_2, z_1, z_2 and π_1 (π_2 is given by the equation $\pi_1 + \pi_2 = 1$). The trivial observation is that the two vectors

$$(\pi_1, x_1, y_1, z_1) \qquad \text{and} \qquad (\pi_2, x_2, y_2, z_2)$$

are interchangeable. The best we can hope for is to solve for them up to permutation. What we study in this article is the inversion problem for a generalisation of the above, where we allow r populations, and on each member of the sample we measure n variables independent of each other. In the example above the variables were binary; this is only to keep the example simple. If, in the example, we allow x_1, x_2, y_1, y_2, z_1 and z_2 to be real valued functions of suitable variables then the general case reduces to the binary one. The reader can find a discussion of more realistic (non-parametric) examples, and some references to the statistical literature, in [7] and Section 1 of this article.

This ends the statistical introduction. Throughout this article k will be a field of characteristic zero. All varieties will be over k. Thus $\mathbb{A}^{n+1}$ means $\mathbb{A}^{n+1}_k$.

Let r and n be positive integers. The symmetric group Σ_r acts on $\{\mathbb{A}^{n+1}\}^r = \mathbb{A}^{(n+1)r}$ by permuting the r copies of $\mathbb{A}^{n+1}$. Let us choose coordinates. We have r copies of $\mathbb{A}^{n+1}$. If i is an integer with $1 \leqslant i \leqslant r$, then the coordinates of the ith copy of $\mathbb{A}^{n+1}$ will be $(x_{i0}, x_{i1}, x_{i2}, \ldots, x_{in})$. Let x stand for the entire $r \times (n+1)$ matrix of coordinates x_{ij}.

Let $\varepsilon = (\varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ be a vector in $\mathbb{N}^{n+1}$. In this paper, $\mathbb{N}$ is the set $\mathbb{N} = \{0, 1, 2, \ldots\}$ of all non-negative integers; thus $0 \in \mathbb{N}$. We can then form the functions

$$\Phi_\varepsilon(x) = \sum_{i=1}^{r} \prod_{j=0}^{n} x_{ij}^{\varepsilon_j}.$$

Now we make the two restrictions.

HYPOTHESIS 0.1. — *The vector $\varepsilon \in \mathbb{N}^{n+1}$ satisfies Hypothesis 0.1 if*

(i) $\varepsilon_0 = 1$,

(ii) $\varepsilon_j \in \{0, 1\}$ for all $1 \leqslant j \leqslant n$.

There are 2^n functions $\Phi_\varepsilon : \mathbb{A}^{(n+1)r} \longrightarrow \mathbb{A}^1$, with ε satisfying Hypothesis 0.1. All of them are symmetric with respect to the action of Σ_r. Combining them, we have a function

$$\Phi : \mathrm{Sym}^r \mathbb{A}^{n+1} \longrightarrow \mathbb{A}^{2^n}.$$

The main result of this article is

THEOREM 0.2. — *For each $r \geqslant 2$ there exists a constant $C(r)$, depending only on r, so that if $n > C(r)$ then the map Φ is birational onto its image.*

Remark 0.3. — It is very easy to see that the map Φ is not biregular onto its image. For example, if $x_{ij} = 0$ for all $j > 1$, then the only non-vanishing Φ_ε satisfying Hypothesis 0.1 are

$$\sum_{i=1}^r x_{i0}, \qquad \sum_{i=1}^r x_{i0}x_{i1}.$$

These two numbers are not enough to allow us to recover the $2r$ unknowns x_{i0} and x_{i1}.

As we said at the beginning, the motivation for the problem comes from mixture models in statistics. If we let $x_{i0} = \pi_i$ be the "mixing probabilities" (the probabilities of a person being Greek or Turkish in the example above), then the Φ_ε, with ε satisfying Hypothesis 0.1, are what is estimable for the sample. The inversion problem is to recover from them the x_{ij}'s (and the π_i's if they are unknown).

In the applications the mixing probabilities $x_{i0} = \pi_i$ are often assumed known. The following result is therefore relevant to the statistical applications:

THEOREM 0.4. — *Assume $n > C(r)$, as in Theorem 0.2. Let $(\pi_1, \ldots, \pi_r)$ be any vector in k^r. Let $\mathbb{A}^{nr} \subset \mathbb{A}^{(n+1)r}$ be the subvariety given by the equations*

$$x_{10} = \pi_1, \qquad x_{20} = \pi_2, \qquad \ldots, \qquad x_{r0} = \pi_r.$$

In other words, we fix the mixing probabilities. The image of the subvariety $\mathbb{A}^{nr} \subset \mathbb{A}^{(n+1)r}$, under the natural projection $\mathbb{A}^{(n+1)r} \longrightarrow \mathrm{Sym}^r \mathbb{A}^{n+1}$, is contained in the exceptional locus of the map Φ if and only if one of the π_i vanishes.

Remark 0.5. — The "if" part of Theorem 0.4 is obvious. For example if $\pi_1 = 0$, then we have no information about x_{1j} with $1 \leqslant j \leqslant n$. What is being asserted is that if all of the π_i are non-zero then the inversion problem is solvable on a large open set.

Now we should say a few words about the number $C(r)$ in Theorems 0.2 and 0.4. The naïve observation is the following. If the map $\Phi : \mathrm{Sym}^r \mathbb{A}^{n+1} \longrightarrow \mathbb{A}^{2^n}$ is to be generically injective, then the dimension of $\mathrm{Sym}^r \mathbb{A}^{n+1}$ must be less than or equal to the dimension of $\mathbb{A}^{2^n}$. That is,

$$(n+1)r \leqslant 2^n.$$

It has been known for many years that this is not enough. In [5, page 221] the reader can find the following example: If $r = 3$ and $n = 4$, then $(n+1)r = 15$ and $2^n = 16$. The dimensions are such that Φ could be generically injective. But it turns out that the image of Φ is 14–dimensional. The map Φ is not even generically finite. Thus it is necessary to take n sufficiently large; there must be a $C(r)$ in the theorems.

Our proof gives an effective upper bound for the optimal $C(r)$. The estimate given by our proof grows asymptotically like $c \cdot r\ln(r)$ where the constant c is effective; in fact, $c \leqslant 6$. The easy lower bound for $C(r)$ is $c \cdot \ln(r)$, with $c \geqslant 1/\ln(2)$. Since $c_1 \cdot \ln(r) \leqslant C(r) \leqslant c_2 \cdot r\ln(r)$ is a fairly big range, it would be interesting to get sharper estimates for the best $C(r)$.

The fact that the problem can be formulated as a question in algebraic geometry is not new. See for example Garcia, Stillman and Sturmfels [3, Section 7]. The formulation given there is very slightly different from our version. We explain.

Consider the Segre embedding $\varphi : \{\mathbb{P}^1\}^n \longrightarrow \mathbb{P}^{2^n-1}$. On $\{\mathbb{A}^1\}^n \subset \{\mathbb{P}^1\}^n$ the map takes the vector $p \in \{\mathbb{A}^1\}^n$, with coordinates $(x_1, x_2, \ldots, x_n)$, to the vector $\varphi(p) \in \mathbb{P}^{2^n-1}$, whose coordinates are $\prod_{i \in I} x_i$ with $I \subset \{1, 2, \ldots, n\}$. Now consider r points $\{p_1, p_2, \ldots, p_r\}$ in $\{\mathbb{A}^1\}^n$. Let the coordinates of p_i be $(x_{i1}, x_{i2}, \ldots, x_{in})$. Then the point

$$\pi_1 \varphi(p_1) + \pi_2 \varphi(p_2) + \cdots + \pi_r \varphi(p_r)$$

is precisely the same as

$$\Phi\Big((\pi_1, p_1), (\pi_2, p_2), \ldots, (\pi_r, p_r) \Big).$$

The closure of the image of the map Φ can therefore be identified with the $(r-1)$–secant variety of the Segre embedding of $\{\mathbb{P}^1\}^n$. The points in the image are the linear combinations of r points in the Segre embedding.

Until this article, all that seems to have been studied systematically is the dimension and the equations of the secant variety. In terms of the map Φ of this article, what has been studied are the equations for the Zariski closure of the image of Φ, and whether the map Φ is generically finite. For the equations of the Zariski closure the reader is referred to Garcia–Stillman–Sturmfels [3] and Landsberg–Manivel [8]. For the dimension of the image of Φ, see Catalisano–Geramita–Gimigliano [1,2]. We should say a little about the methods of this last article.

To compute the dimension of the image of Φ Catalisano, Geramita and Gimigliano make the following observation. For a general point $x \in \mathrm{Sym}^r\mathbb{A}^{n+1}$, the derivative of Φ at x maps the tangent space at x onto the tangent space at $\Phi(x)$ to the variety $\Phi(\mathrm{Sym}^r\mathbb{A}^{n+1}) \subset \mathbb{A}^{2^n}$. It is easy enough to express the dimension of the image tangent space in terms of some linear system on $\{\mathbb{P}^1\}^n$. The problem then becomes to estimate the dimension of this linear system.

The reader can, of course, find more detail in [1]. The reason for our brief discussion of the methods is only to explain that our result cannot possibly be obtained by these techniques. A tangent space computation will at most establish that the map Φ is generically finite; it cannot hope to show that it is generically injective.

In Sections 2 and 3 we set up the notation and prove some elementary facts about the rings concerned. The notation attempts to be consistent with Neeman [13]. The wrinkle is that it is very convenient, for our current problem, to pass to the inverse limit as $r \longrightarrow \infty$ of the rings involved. Section 3 develops the elementary properties of this inverse limit ring. The proofs of the main theorems are in Section 4.

The authors would like to thank the referee for many very helpful suggestions for improvements.

1. Statistical motivation.

Suppose a population consists of r different sub-populations, and that the sampled data from each sub-population are vectors of length n. In one

of many examples which motivate statisticians to study this problem, the "full" or "mixture" population is the population of people who visit a doctor, concerned that they might be suffering a certain illness; $r = 2$; and the two sub-populations represent people who suffer the illness, and those who do not have the illness, respectively. For each presenting patient the doctor records a vector of n continuous measurements. However, only by invasive surgery, which is both expensive and potentially dangerous, can the doctor determine for certain whether the patient is ill or not.

Thus, we can never be certain whether a given patient actually has the illness. However, it is nevertheless of practical interest to know the distribution of symptom vectors for the sub-population of ill patients, and for the sub-population of well patients; and it is also desirable to know the percentage of patients in the population who are actually ill, even though these patients cannot be identified individually. How much structure must we assume in order to be able to accurately estimate these unknowns from data?

More generally, in the case of r sub-populations, a statistician would consider the ith population to be present in proportion π_i, where $\pi_1 + \ldots + \pi_r = 1$; and would assume that a data vector from the mixture population was obtained by first choosing a sub-population according to a sampling scheme which assigned probability π_i to the ith sub-population. This is termed a "mixture model," the probabilities π_i are called the "mixing proportions," and the full population is called the "mixture population." In this general setting of r sub-populations of vectors of length n, a statistical problem of significant interest is: how much must we assume about the sub-populations in order to be able to estimate the distributions of individual sub-populations, and to estimate the so-called mixing proportions, given only data from the mixture population?

One way of tackling this problem is to assume that the distribution of each sub-population is the product of its marginal distributions — that is, within each sub-population the components of the n-vector of data are statistically independent. Under this constraint the problem actually becomes increasingly restricted as dimension increases, and one might conjecture that for each r there is a minimal n, $C(r)$ say, such that one can generically estimate all the mixture proportions π_i, and all the sub-population distributions, given enough data on the mixture distribution, provided $n \geqslant C(r)$. One does not need to make any assumptions about

the nature of the marginal distributions. A statistical interpretation of Theorem 0.2 is that there does in fact exist such a $C(r)$.

It is the generality of this solution which makes it of particular interest to statisticians. In more conventional, "parametric" mixture problems it is assumed that the sub-population distributions are determined by a finite number of parameters. (For example, the parameters might be the means and variances of normal distributions.) Attention then focuses on estimating the parameters. That is often a relatively standard problem, which can be solved quite simply. The setting of the present paper is "nonparametric," however, in the sense that no assumptions are made about the sub-population distribution. The beauty of the solution given by Theorem 0.2 is that it shows that regardless of the number of parameters (including the case where they are infinite in number), the sub-population distributions can be identified from data on the mixture population, provided only that $n \geqslant C(r)$.

The problem of recovering the mixing proportions and sub-population distributions is known as an identifiability problem in the statistical literature. Teicher [15], [16] and Yakowitz and Spragins [19] provide general theories on the identifiability in finite mixture models; however, their work is specific to mixtures of certain classes of parametric distribution functions. More recent, Hall and Zhou [6] propose a nonparametric method of inversion in the case $r = 2$. For detailed accounts of mixture models, we refer the interested reader to Titterington, Smith, and Makov [17], Lindsay [9], and MacLachlan and Peel [10].

2. Notational conventions and background.

As in the introduction the affine space $\mathbb{A}^{(n+1)r}$ is given the coordinates x_{ij}, with $1 \leqslant i \leqslant r$ and $0 \leqslant j \leqslant n$. We let x stand for the $r \times (n+1)$ matrix $x = (x_{ij})$ of all the coordinates. Let $S_r = k[x]$ be the ring of all polynomials in the variables (x_{ij}). Then $\mathbb{A}^{(n+1)r} = \mathrm{Spec}(S_r)$. Let $S_r^{\Sigma_r} \subset S_r$ be the subring of Σ_r–invariant polynomials. Almost by definition, $\mathrm{Sym}^r \mathbb{A}^{n+1} = \mathrm{Spec}(S_r^{\Sigma_r})$.

Remark 2.1. — We remind the reader: This means that the polynomial functions on the symmetric power $\mathrm{Sym}^r \mathbb{A}^{n+1}$ are precisely the polynomials on $\mathbb{A}^{(n+1)r}$ invariant under the action of Σ_r.

Now we want to establish our notation for elements of the ring S_r and its subring $S_r^{\Sigma_r}$. Monomials in the variables $\{x_{ij} \mid i \text{ fixed}, 0 \leqslant j \leqslant n\}$ will be denoted y_i or z_i. In other words, y_i is an expression

$$x_{i0}^{\varepsilon_0} x_{i1}^{\varepsilon_1} \cdots x_{in}^{\varepsilon_n}.$$

By y_j we will mean the same monomial, but where i and j are interchanged. That is, y_j is the monomial

$$x_{j0}^{\varepsilon_0} x_{j1}^{\varepsilon_1} \cdots x_{jn}^{\varepsilon_n}.$$

If we have more than two such monomials (that is, the letters y and z no longer suffice) the notation will be $\{y_{i1}, y_{i2}, \ldots, y_{it}\}$. Thus each $y_{i\ell}$ is a monomial in the x_{ij} for i fixed. The first subscript of $y_{i\ell}$ tells us which variables x_{ij} are permissible. And once again, $y_{i\ell}$ and $y_{j\ell}$ are the same monomial, up to interchanging i and j. The monomials in the ring $S_r = k[x]$ are of the form

$$y_{11} y_{22} \cdots y_{rr}.$$

The group Σ_r permutes the monomials.

DEFINITION 2.2. — *Let $y_{11} y_{22} \cdots y_{rr}$ be a monomial in $S_r = k[x]$. Then*

$$m(y_{11} y_{22} \cdots y_{rr}).$$

is defined to be the sum of all the monomials in the Σ_r–orbit of $y_{11} y_{22} \cdots y_{rr}$. Clearly, $m(y_{11} y_{22} \cdots y_{rr})$ is an element of $S_r^{\Sigma_r} \subset S_r$.

DEFINITION 2.3. — *The polynomials $m(y_{11} y_{22} \cdots y_{rr})$ will be called the molecules of $S_r^{\Sigma_r}$. If y_{11} is non-constant and $y_{22} = y_{33} = \cdots = y_{rr} = 1$, then the polynomial*

$$m(y_{11}) = m(y_{11} y_{22} \cdots y_{rr})$$

will be called an atom of $S_r^{\Sigma_r}$.

Remark 2.4. — The molecules form a basis for $S_r^{\Sigma_r}$, as a vector space over k.

Remark 2.5. — If $\varepsilon = (\varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is a non-zero vector in $\mathbb{N}^{n+1}$ and

$$y_1 = x_{10}^{\varepsilon_0} x_{11}^{\varepsilon_1} \cdots x_{1n}^{\varepsilon_n},$$

then the atom $m(y_1)$ is nothing other than the polynomial $\Phi_\varepsilon(x)$ of the introduction.

DEFINITION 2.6. — *The atom* $m(y_1) = \Phi_\varepsilon(x)$ *is called*

2.6.1 Special *if, for all* $0 \leqslant j \leqslant n$, $\varepsilon_j \in \{0, 1\}$;

2.6.2 Extra Special *if it is special, and furthermore* $\varepsilon_0 = 1$.

Remark 2.7. — The extra special atoms $m(y_1) = \Phi_\varepsilon(x)$ are precisely the ones for which ε satisfies Hypothesis 0.1. Theorem 0.2 asserts that the 2^n extra special atoms $m(y_1)$ define a birational map from $\mathrm{Sym}^r \mathbb{A}^{n+1}$ onto its image. Translated to commutative algebra, this means that the 2^n extra special atoms generate the quotient field of $S_r^{\Sigma_r}$. Every element of $S_r^{\Sigma_r}$ can be written as a quotient of polynomials in the extra special $m(y_1)$. Theorem 0.4 asserts further that the subvariety $\mathbb{A}^{nr}$, given by the equations

$$x_{10} = \pi_1, \qquad x_{20} = \pi_2, \qquad \ldots, \qquad x_{r0} = \pi_r,$$

will lie in the exceptional locus if and only if one of the π_i vanishes. Translated to commutative algebra this means the following. Assume all of the π_i are non-zero. By Theorem 0.2 every element of $S_r^{\Sigma_r}$ can be written as a quotient of polynomials in the extra special atoms. It is possible to do so in such a way that the denominators do not vanish identically on the subvariety $\mathbb{A}^{nr} \subset \mathbb{A}^{(n+1)r}$.

In the proof we will first show that the special atoms generate the quotient field of $S_r^{\Sigma_r}$ (Theorem 4.8), and then that it is possible to write every special atom as a quotient of polynomials in extra special atoms (Theorem 4.15). The statement about the denominators will be obvious from the constructions.

For every $r \geqslant 2$ we have a ring homomorphism $S_r \longrightarrow S_{r-1}$. It simply sends all the variables x_{rj} to zero. This homomorphism takes $S_r^{\Sigma_r} \subset S_r$ into $S_{r-1}^{\Sigma_{r-1}} \subset S_{r-1}$. Even better, for any $\ell < r$ the homomorphism takes

$$m(y_{11}y_{22}\cdots y_{\ell\ell}) \in S_r^{\Sigma_r} \quad \text{to} \quad m(y_{11}y_{22}\cdots y_{\ell\ell}) \in S_{r-1}^{\Sigma_{r-1}}.$$

DEFINITION 2.8. — *Let* $\varprojlim S_r^{\Sigma_r}$ *be the inverse limit of the rings* $S_r^{\Sigma_r}$. *Let* $y_{11}y_{22}\cdots y_{\ell\ell}$ *be a monomial in the ring* S_ℓ. *For each* $r \geqslant \ell$ *the ring* $S_r^{\Sigma_r}$ *contains the molecule* $m(y_{11}y_{22}\cdots y_{\ell\ell})$, *and these molecules are compatible*

with the structure homomorphisms of the inverse system for $r > \ell$. In an abuse of notation, we let $m(y_{11}y_{22}\cdots y_{\ell\ell}) \in \varprojlim S_r^{\Sigma_r}$ stand for the inverse limit. Let S be the vector subspace of $\varprojlim S_r^{\Sigma_r}$ spanned by all the $m(y_{11}y_{22}\cdots y_{\ell\ell})$. As with $S_r^{\Sigma_r}$, the molecules of S are the $m(y_{11}y_{22}\cdots y_{\ell\ell})$, and the atoms are the $m(y_{11})$ with y_{11} non-constant.

Remark 2.9. — The vector subspace $S \subset \varprojlim S_r^{\Sigma_r}$ is a subring. The product of two molecules can easily be expanded as a linear combination of molecules. Furthermore, the molecules are linearly independent; they form a basis for S over k.

We will need the following, useful definition:

Definition 2.10. — Suppose y_{11}, y_{22}, ..., $y_{\ell\ell}$ are all non-constant monomials. Then we will say that the molecule $m(y_{11}y_{22}\cdots y_{\ell\ell})$ has length ℓ. The degree of the molecule $m(y_{11}y_{22}\cdots y_{\ell\ell})$ is the total degree in the variables x_{ij}.

Remark 2.11. — The natural map $S \subset \varprojlim S_r^{\Sigma_r} \longrightarrow S_r^{\Sigma_r}$ is obviously surjective. The kernel is spanned by the molecules $m(y_{11}y_{22}\cdots y_{\ell\ell})$ of length $\ell > r$.

3. Preliminary results.

Lemma 3.1. — Let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_\ell$ be integers > 0, and let y_{11}, y_{12}, ..., $y_{1\ell}$ be distinct non-constant monomials in the variables $\{x_{10}, x_{11}, x_{12}, \ldots, x_{1n}\}$. By Remark 2.4, the product of atoms

$$m(y_{11})^{\varepsilon_1} m(y_{12})^{\varepsilon_2} \cdots m(y_{1\ell})^{\varepsilon_\ell}$$

can be expanded uniquely as a linear combination of molecules (they form a basis for S). Then in this expansion there is exactly one molecule of length

$$\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_\ell.$$

All the other molecules are shorter. Furthermore, from this long molecule we can uniquely recover all the pairs (y_{1j}, ε_j), up to permutation.

Proof. — The unique long molecule of length $\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_\ell$ is

$$m\big(y_{11}y_{21}\cdots y_{\varepsilon_1 1}y_{(\varepsilon_1+1)2}\cdots y_{(\varepsilon_1+\varepsilon_2)2}\cdots y_{(\varepsilon_1+\cdots+\varepsilon_{\ell-1}+1)\ell}\cdots y_{(\varepsilon_1+\cdots+\varepsilon_\ell)\ell)}\big)$$

and it is clear that from it we can recover all the pairs (y_{1j}, ε_j).   □

LEMMA 3.2. — *The atoms of S are algebraically independent over the field k.*

Proof. — We prove this by contradiction. Suppose we have a non-trivial algebraic identity. Such an identity is a polynomial $P(Y_1, \ldots, Y_\ell)$, which vanishes on some distinct set of atoms $\{Y_1, \ldots, Y_\ell\}$. That is, there exist distinct non-constant monomials $\{y_{1j},\ 1 \leqslant j \leqslant \ell\}$ in $\{x_{10}, x_{11}, x_{12}, \ldots, x_{1n}\}$, so that if we substitute

$$Y_1 = m(y_{11}), \qquad Y_2 = m(y_{12}), \qquad \ldots, \qquad Y_\ell = m(y_{1\ell})$$

then $P(Y_1, \ldots, Y_\ell) = 0$. Now consider the terms of highest degree M. They are terms

$$\lambda Y_1^{\varepsilon_1} Y_2^{\varepsilon_2} \cdots Y_\ell^{\varepsilon_\ell} \quad \text{with} \quad \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_\ell = M.$$

By Lemma 3.1, if we substitute $Y_j = m(y_{1j})$ and expand, each of these monomials $Y_1^{\varepsilon_1} Y_2^{\varepsilon_2} \cdots Y_\ell^{\varepsilon_\ell}$ will give exactly one molecule of length M, and these molecules will all be distinct. There can be no cancellation.   □

LEMMA 3.3. — *Suppose $y_{11}, y_{12}, \ldots, y_{1\ell}$ are non-constant monomials in the variables $\{x_{10}, x_{11}, x_{12}, \ldots, x_{1n}\}$. Then the molecule $m(y_{11}y_{22}\cdots y_{\ell\ell})$ $\in S$ is equal, modulo products of molecules of lower degree and shorter length, to*

$$\lambda m(y_{11}y_{12}\cdots y_{1\ell}).$$

Here λ is a rational number with $(-1)^{\ell-1}\lambda > 0$, in particular $\lambda \neq 0$. Furthermore the molecules of lower degree and shorter length, whose products occur in the expansion, may each be written in the form

$$m(z_{11}z_{22}\cdots z_{tt}).$$

For any subset $B \subset \{1, 2, \ldots, t\}$ there exists a subset $A \subset \{1, 2, \ldots, \ell\}$ and an identity

$$\prod_{j\in B} z_{1j} = \prod_{j\in A} y_{1j}.$$

Proof (cf. [13, proof of Lemma 2]). — We prove the statement by induction on the length ℓ. Suppose therefore that it is known for all molecules $m(z_{11}z_{22}\cdots z_{tt})$ of length $< \ell$. Now

$$m(y_{11}y_{22}\cdots y_{\ell\ell}) = \alpha_1 \cdot m(y_{11}) \cdot m(y_{22}\cdots y_{\ell\ell}) - \sum_{j=2}^{\ell} \alpha_j \cdot m(y_{22}\cdots z_{jj}\cdots y_{\ell\ell})$$

where $z_{jj} = y_{j1}y_{jj}$, and α_j, $1 \leqslant j \leqslant \ell$ are all positive rational numbers. Since all the molecules on the right hand side are of length $< \ell$, we can apply induction.   □

DEFINITION 3.4. — *Suppose y_{11}, y_{12}, ..., $y_{1\ell}$ are monomials in the variables $\{x_{10}, x_{11}, x_{12}, \ldots, x_{1n}\}$. For every subset $A \subset \{1, 2, \ldots, \ell\}$, we can form the monomial $z_{1A} = \prod_{j\in A} y_{1j}$. The monomials y_{11}, y_{12}, ..., $y_{1\ell}$ are in general position if the 2^ℓ monomials z_{1A} are all distinct.*

Remark 3.5. — If y_{11}, y_{12}, ..., $y_{1\ell}$ are in general position, then the formula of the proof of Lemma 3.3

$$m(y_{11}y_{22}\cdots y_{\ell\ell}) = \alpha_1 \cdot m(y_{11}) \cdot m(y_{22}\cdots y_{\ell\ell}) - \sum_{j=2}^{\ell} \alpha_j \cdot m(y_{22}\cdots z_{jj}\cdots y_{\ell\ell})$$

simplifies; we have $\alpha_1 = \alpha_2 = \cdots = \alpha_\ell = 1$. The statement of the lemma becomes more precise: the value of λ is easily computed, by induction, to be $(-1)^{\ell-1}(\ell-1)!$

Since the above is a formal identity, the reader may wonder why we cannot substitute $y_{11} = y_{12}$. The problem is that the rational coefficients change. Let us do an example in $S_2^{\Sigma_2}$. The molecule $m(y_{11}y_{22})$ is

$$m(y_{11}y_{22}) = y_{11}y_{22} + y_{21}y_{12}.$$

If we substitute $y_{11} = y_{12}$ we get

$$y_{11}y_{21} + y_{21}y_{11} = 2y_{11}y_{21}$$

and this is $2m(y_{11}y_{21})$. In other words, it is OK to substitute if we are careful about the coefficients.

Because it is easier to do the computations assuming the y_{11}, y_{12}, ..., $y_{1\ell}$ are in general position, we will permit ourselves to do so. When necessary we will make substitutions, and keep track of what happens to the rational coefficients.

PROPOSITION 3.6. — *Every molecule $m(y_{11}y_{22}\cdots y_{\ell\ell})$ can be uniquely expanded in S as a polynomial in the atoms. The atoms that occur in this expansion are all of the form $m(z_{1A})$, where $z_{1A} = \prod_{j\in A} y_{1j}$ for some subset $A \subset \{1, 2, \ldots, \ell\}$.*

Proof. — By Lemma 3.2 we know that the atoms are algebraically independent; if a molecule has an expansion as a polynomial in the atoms, that expansion must be unique. We only need to prove that the expansion exists, and that the atoms that occur are of the right sort. We prove this by induction on the length, the case of molecules of length 1 being obvious.

Suppose therefore that we know it whenever the length is $< \ell$. We need to show that $m(y_{11}y_{22}\cdots y_{\ell\ell})$ is in the algebra generated by the atoms of the form $m(z_{1A})$. Lemma 3.3 tells us that $m(y_{11}y_{22}\cdots y_{\ell\ell})$ is a sum of products of molecules of shorter length. [It also tells us that the coefficient of the molecule of length 1 is non-zero, which is irrelevant here.] Induction applies. $\square$

COROLLARY 3.7. — *The ring S is a polynomial ring on the atoms.*

One easily deduces:

PROPOSITION 3.8. — *The ring $S_r^{\Sigma_r}$ is generated, as an algebra over k, by the atoms of degree $\leqslant r$.*

Proof. — The homomorphism $S \longrightarrow S_r^{\Sigma_r}$ is surjective, and by Proposition 3.6 (or Corollary 3.7) the ring S is generated as an algebra by the atoms. Hence so is $S_r^{\Sigma_r}$. The problem is to show that, in $S_r^{\Sigma_r}$, any atom of degree $> r$ can be expressed as a polynomial in atoms of lower degrees.

Let y_1 be a monomial in the variables $\{x_{10}, x_{11}, x_{12}, \ldots, x_{1n}\}$ of degree $s > r$. Then we can choose some function

$$f : \{1, 2, 3, \ldots, s\} \longrightarrow \{0, 1, 2, \ldots, n\}$$

so that

$$y_1 = x_{1f(1)}x_{1f(2)}x_{1f(3)} \cdots x_{1f(s)}.$$

Now we apply Lemma 3.3 to the molecule

$$m(x_{1f(1)}x_{2f(2)}x_{3f(3)}\cdots x_{sf(s)}).$$

Since the molecule has length $s > r$, it lies in the kernel of the homomorphism $S \longrightarrow S_r^{\Sigma_r}$. Lemma 3.3 tells that this molecule is equal in S, modulo products of molecules of smaller degree, to $\lambda m(y_1)$, with $\lambda \neq 0$. In $S_r^{\Sigma_r}$ we deduce that $m(y_1)$ is equal to a sum of products of molecules of smaller degree. Expanding these molecules into their atoms as in Proposition 3.6, we conclude that in the ring $S_r^{\Sigma_r}$ the atom $m(y_1)$ can be written as a polynomial in atoms of degree $< s$. □

Remark 3.9. — Proposition 3.8, giving generators for the ring $S_r^{\Sigma_r}$, is not new. It may essentially be found in Weyl [18, Chapter II, §3]. What seems to be new is Corollary 3.7, asserting that the inverse limit ring S is a polynomial algebra on the atoms.

Since we are interested in the problem only up to birational maps, a set of generators for the quotient field suffices. There is old literature about generators for the quotient field. The old theorem asserts that the field is purely transcendental; it has exactly $r(n+1)$ generators. The theorem was known in the late 19th century; see Netto [14, Bd.2, §383]. The theorem was forgotten and then reproved by Mattuck [11]. See also Gelfand, Kapranov and Zelevinsky [4, page 137, Theorem 2.8'].

The generators given in the old theorem were not atoms, but we have seen in this section that the atoms generate the ring $S_r^{\Sigma_r}$. Rewriting the old theorem in terms of atoms, we have:

THEOREM 3.10. — *The $r(n+1)$ atoms $m(x_{10}^{\ell}x_{1j})$, with $0 \leqslant \ell \leqslant r-1$ and $0 \leqslant j \leqslant n$, generate the quotient field of $S_r^{\Sigma_r}$.*

Since these atoms are not special they do not help us much in this article. The interested reader can substitute, in the first line of the proof of Theorem 4.8, the generators given by Theorem 3.10 in place of those given by Proposition 3.8. It gives a slight improvement for the integer $C(r)$ of the introduction, but at the cost of making the exceptional locus worse.

The authors want to thank Brion, Drinfeld and Larry Smith for helpful pointers to the classical literature.

4. Proofs of the main theorems.

Proposition 3.6 tells us that every molecule can be uniquely expanded in S as a polynomial in the atoms. It is natural to ask what is the coefficient of an atom $Y = m(y_1)$ in the expansion. We better make precise what we mean by this. The molecule $m(y_{11}y_{22}\cdots y_{\ell\ell})$ is equal to a unique polynomial $P(Y_1, Y_2, \ldots, Y_m)$, with Y_i atoms. There is a unique way to write

$$P(Y_1, Y_2, \ldots, Y_m) \;=\; Y_1 Q(Y_1, Y_2, \ldots, Y_m) \;+\; R(Y_2, \ldots, Y_m)$$

where $R(Y_2, \ldots, Y_m)$ is a polynomial only in the atoms $\{Y_2, \ldots, Y_m\}$. By the "coefficient of Y_1" we understand the polynomial $Q(Y_1, Y_2, \ldots, Y_m)$ in this expression. Can we compute it? Since S is the polynomial ring in the atoms, any description of $Q(Y_1, Y_2, \ldots, Y_m)$ as an element of S will suffice.

The answer is that there is a relatively simple expression for it, as long as the monomials occuring in the molecule are in general position. We remind the reader of Definition 3.4: The monomials y_{11}, y_{12}, ..., $y_{1\ell}$ in the variables $\{x_{10}, x_{11}, x_{12}, \ldots, x_{1n}\}$ are in general position if the 2^ℓ monomials $z_{1A} = \prod_{j \in A} y_{1j}$ are all distinct.

PROPOSITION 4.1. — *Let y_{11}, y_{12}, ..., $y_{1\ell}$ be monomials in general position. Let s be an integer, with $1 \leqslant s \leqslant \ell$. The coefficient of the atom $m(y_{11}y_{12}\cdots y_{1s})$ in the expansion of the molecule $m(y_{11}y_{22}\cdots y_{\ell\ell})$ is*

$$(-1)^{s-1}(s-1)! \, m(y_{(s+1)(s+1)}\cdots y_{\ell\ell}).$$

Proof. — We observe the identity in S

$$m(y_{11}y_{22}\cdots y_{\ell\ell}) = m(y_{11}y_{22}\cdots y_{ss}) \cdot m(y_{(s+1)(s+1)}\cdots y_{\ell\ell})$$

$$- \sum m(z_{11}z_{22}\cdots z_{tt}).$$

In the molecules $m(z_{11}z_{22}\cdots z_{tt})$ that occur in the sum, the monomial z_{11} may be chosen to be of the form $y_{1\alpha}y_{1\alpha'}$, with $1 \leqslant \alpha \leqslant s$ and $s+1 \leqslant \alpha' \leqslant \ell$. Proposition 3.6 states that all the atoms in the expansion of $m(z_{11}z_{22}\cdots z_{tt})$ are of the form $m(\prod_{j \in A} z_{1j})$, for some $A \subset \{1, 2, \ldots, t\}$. If $1 \in A$ then $\prod_{j \in A} z_{1j}$ is divisible by $y_{1\alpha'}$, and if $1 \notin A$ then $\prod_{j \in A} z_{1j}$ is not divisible by $y_{1\alpha}$. Since we are assuming the monomials are in general position, $\prod_{j \in A} z_{1j}$ cannot equal $y_{11}y_{12}\cdots y_{1s}$.

We conclude that the coefficient of the atom $Y = m(y_{11}y_{12}\cdots y_{1s})$ in the expansion of the molecule $m(y_{11}y_{22}\cdots y_{\ell\ell})$ is equal to the coeficient of Y in the expansion of $m(y_{11}y_{22}\cdots y_{ss})\cdot m(y_{(s+1)(s+1)}\cdots y_{\ell\ell})$. By Remark 3.5 the coefficient of $m(y_{11}y_{12}\cdots y_{1s})$ in $m(y_{11}y_{22}\cdots y_{ss})$ is $(-1)^{s-1}(s-1)!$, and the proposition follows immediately. $\hfill\square$

Notation. — In the next couple of lemmas we will fix $r+2$ monomials in general position $\{a_1, b_1, y_{13}, y_{14}, \ldots, y_{1(r+2)}\}$, in the variables $\{x_{10}, x_{11}, x_{12}, \ldots, x_{1n}\}$. We let $R \subset S$ be the subalgebra generated by atoms of the form

$$m(a_1^{\varepsilon_1} b_1^{\varepsilon_2} y_{13}^{\varepsilon_3} y_{14}^{\varepsilon_4} \cdots y_{1(r+2)}^{\varepsilon_{r+2}})$$

where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{r+2}$ all lie in the set $\{0, 1\}$, and $\varepsilon_1 + \varepsilon_2 \leqslant 1$. Given two elements $\varphi, \theta \in S$, we say that φ and θ are *congruent* if $\varphi - \theta$ lies in R. In symbols, we will denote this relation $\varphi \equiv \theta$.

LEMMA 4.3. — *With the notation as in 4.2, the expression*

$$m(a_1 b_1 y_{33} y_{44} \cdots y_{(r+2)(r+2)}) \ - \ \sum_{j=3}^{r+2} m(a_1 b_2 y_{2j} y_{33} \cdots \widehat{y_{jj}} \cdots y_{(r+2)(r+2)})$$

is congruent to an expression

$$m(a_1 b_1) \cdot m(y_{33} y_{44} \cdots y_{(r+2)(r+2)}) - \sum \alpha_I \cdot m(a_1 b_1 z_{11}) \cdot m(z_{22} z_{33} \cdots z_{\ell\ell})$$

where the α_I are rational numbers, the I in the subscript runs over the partitions of the set $\{3, 4, \ldots, r+2\}$ into $\ell < r$ sets, each monomials z_{1j} is equal to some $\prod_{j \in A} y_{1j}$ with $A \subset \{3, 4, \ldots, r+2\}$, and

$$\prod_{j=1}^{\ell} z_{1j} = \prod_{j=3}^{r+2} y_{1j}.$$

Proof. — Consider the molecule $m(a_1 b_1 y_{33} y_{44} \cdots y_{(r+2)(r+2)})$. We may expand it as a polynomial in its atoms. The expansion cannot have a term divisible by $m(a_1 b_1 z_{11}) \cdot m(a_1 b_1 z_{12})$. Modulo elements of the subring $R \subset S$, the molecule $m(a_1 b_1 y_{33} y_{44} \cdots y_{(r+2)(r+2)})$ must be of the form

$$\sum r_i \cdot m(a_1 b_1 z_{1i}),$$

with $r_i \in R$. But then Proposition 4.1 computed for us the coefficient r_i. We have that $m(a_1 b_1 y_{33} y_{44} \cdots y_{(r+2)(r+2)})$ is congruent to

$$m(a_1 b_1) \cdot m(y_{33} y_{44} \cdots y_{(r+2)(r+2)})$$

$$- \sum_{j=3}^{r+2} m(a_1 b_1 y_{1j}) \cdot m(y_{33} \cdots \widehat{y_{jj}} \cdots y_{(r+2)(r+2)})$$

$$+ \sum \alpha_I \cdot m(a_1 b_1 z_{11}) \cdot m(z_{22} z_{33} \cdots z_{\ell\ell})$$

where $z_{11} = \prod_{j \in A} y_{1j}$, the set $A \subset \{3, 4, \ldots, r+2\}$ has at least two elements, and $z_{12}, z_{13}, \ldots, z_{1\ell}$ are some permutation of y_{1t}, $t \notin A$. In particular $\ell < r$.

Similarly, the molecule $m(a_1 b_2 y_{2j} y_{33} \cdots \widehat{y_{jj}} \cdots y_{(r+2)(r+2)})$ is easily seen to be congruent to

$$- m(a_1 b_1 y_{1j}) \cdot m(y_{33} \cdots \widehat{y_{jj}} \cdots y_{(r+2)(r+2)})$$

$$+ \sum \alpha_I \cdot m(a_1 b_1 z_{11}) \cdot m(z_{22} z_{33} \cdots z_{\ell\ell})$$

where $z_{11} = \prod_{j \in A} y_{1j}$, the set $A \subset \{3, 4, \ldots, r+2\}$ has at least two elements, one of which must be j, and $z_{12}, z_{13}, \ldots, z_{1\ell}$ are some permutation of y_{1t}, $t \notin A$. Once again $\ell < r$.

The lemma simply combines these congruences.   □

COROLLARY 4.4. — *In the ring $S_r^{\Sigma_r}$ there is a congruence, modulo the image of $R \subset S$ under the map $S \longrightarrow S_r^{\Sigma_r}$,*

$$m(a_1 b_1) \cdot m(y_{33} y_{44} \cdots y_{(r+2)(r+2)}) \quad \equiv \quad \sum \alpha_I \cdot m(a_1 b_1 z_{11}) \cdot m(z_{22} z_{33} \cdots z_{\ell\ell})$$

where the terms on the right all have $\ell < r$ and

$$\prod_{j=1}^{\ell} z_{1j} = \prod_{j=3}^{r+2} y_{1j}.$$

Proof. — All the molecules in the expression

$$m(a_1 b_1 y_{33} y_{44} \cdots y_{(r+2)(r+2)}) \; - \; \sum_{j=3}^{r+2} m(a_1 b_2 y_{2j} y_{33} \cdots \widehat{y_{jj}} \cdots y_{(r+2)(r+2)})$$

have length $r + 1$, and hence lie in the kernel of $S \longrightarrow S_r^{\Sigma_r}$. But by Lemma 4.3 the expression is congruent modulo $R \subset S$ to

$$m(a_1 b_1) \cdot m(y_{33} y_{44} \cdots y_{(r+2)(r+2)}) - \sum \alpha_I \cdot m(a_1 b_1 z_{11}) \cdot m(z_{22} z_{33} \cdots z_{\ell\ell}).$$

Hence the corollary.   □

Notation 4.5. — In the next lemma we wish to apply Corollary 4.4. Let N be a positive integer, to be determined later. The situation will be as follows. We will be given a_1 and b_1, which are monomials in the variables $x_{10}, \ldots, x_{1p}$. We will assume n is large enough so that we have variables $x_{1(p+1)}, \ldots, x_{1(p+N)}$ (that is, $n \geqslant p + N$). We will partition the variables $x_{1(p+1)}, \ldots, x_{1(p+N)}$ into r non-empty sets, which we will call $A_3, A_4, \ldots, A_{r+2}$. For any integer j with $3 \leqslant j \leqslant r + 2$, put $y_{1j} = \prod_{\alpha \in A_j} x_{1\alpha}$. For each partition, Corollary 4.4 gives us a congruence modulo an R, and the subring $R \subset S$ depends on the partition. The idea is to combine these congruences.

To do this we define a subring $\widetilde{R} \subset S$ which contains all the R's, for every possible partition. The ring $\widetilde{R}$ is the subalgebra of S generated by all the atoms

$$m(a_1^{\varepsilon_1} b_1^{\varepsilon_2} x_{1(p+1)}^{\varepsilon_{p+1}} x_{1(p+2)}^{\varepsilon_{p+2}} \cdots x_{1(p+N)}^{\varepsilon_{p+N}})$$

with $\varepsilon_j \in \{0, 1\}$ for all j, and $\varepsilon_1 + \varepsilon_2 \leqslant 1$.

LEMMA 4.6. — *Let r be an integer, with $r \geqslant 2$. There exists a number $N = N(r)$ with the following property. Let a_1 and b_1 be monomials in the variables $x_{10}, \ldots, x_{1p}$, and assume that n is large enough so that we have variables $x_{1(p+1)}, \ldots, x_{1(p+N)}$ (that is, $n \geqslant p + N$). Then in the ring $S_r^{\Sigma_r}$ there is an identity*

$$s \cdot m(a_1 b_1) \; = \; t,$$

with $s \neq 0$. Both s and t lie in the subalgebra $\widetilde{R} \subset S$ of Notation 4.5, and furthermore s is a polynomial only in the variables $x_{1(p+1)}, \ldots, x_{1(p+N)}$.

Proof. — As in Notation 4.5 we partition the variables $x_{1(p+1)}, \ldots, x_{1(p+N)}$ into r non-empty sets, which we call $A_3, A_4, \ldots, A_{r+2}$. For any

integer j with $3 \leqslant j \leqslant r+2$, put $y_{1j} = \prod_{\alpha \in A_j} x_{1\alpha}$. Corollary 4.4 gives us a congruence modulo $\widetilde{R}$

$$m(a_1 b_1) \cdot m(y_{33} y_{44} \cdots y_{(r+2)(r+2)}) \equiv \sum \alpha_I \cdot m(a_1 b_1 z_{11}) \cdot m(z_{22} z_{33} \cdots z_{\ell\ell}).$$

Now we vary the partitions. For every partition we obtain such a congruence. The idea is to take a linear combination of these congruences so that the right hand side vanishes. Let us first show that this suffices.

Note that the molecules $m(y_{33} y_{44} \cdots y_{(r+2)(r+2)})$ are of length exactly r. They do not vanish in $S_r^{\Sigma_r}$. What is more, as we vary the partitions they yield linearly independent elements of $S_r^{\Sigma_r}$. The left hand side of any non-trivial linear combination of the congruences would be $s \cdot m(a_1 b_1)$ for some $s \neq 0$, s a polynomial in atoms involving only the variables $x_{1(p+1)}, \ldots, x_{1(p+N)}$, and of degree at most one in each.

To say that the linear combination is congruent to zero it to say that it lies in $\widetilde{R}$. Thus the lemma would follow. All we need to prove is that, for some non-trivial linear combination of the congruences, the right hand side vanishes. The point is a very simple counting argument; we will show that, if N is large enough, there are more congruences than the dimension of the vector space of all possible right hand sides.

The number of congruences is simple. It is the number of ways of partitioning a set of N variables into r non-empty slots. Let us label the slots. Then each variable can go into r slots, giving a first estimate of r^N partitions. But in this we counted the partitions in which some slots are empty. There are r slots which could be empty. Therefore $r^N - r(r-1)^N$ is an underestimate of the number of labeled partitions (we have removed the partitions where two or more slots are empty too many times). Anyway, we have at least

$$\frac{r^N - r(r-1)^N}{r!}$$

different congruences.

In each congruence, the terms on the right are linear combinations of expressions

$$m(a_1 b_1 z_{11}) \cdot m(z_{22} z_{33} \cdots z_{\ell\ell})$$

with $\ell < r$. The number of such expressions is the number of ways of partitioning the variables $x_{1(p+1)}, \ldots, x_{1(p+N)}$ into $\ell < r$ slots, with the first one (z_{11}) labeled, but the rest not. For each ℓ we get that $\ell^N/(l-1)!$ is an overestimate; it includes partitions where some of the slots are empty.

Thus the right hand side of each congruence is a linear combination of fewer than

$$\sum_{\ell=1}^{r-1} \frac{\ell^N}{(\ell-1)!}$$

possible polynomials.

As soon as N is chosen large enough so that

$$\frac{r^N - r(r-1)^N}{r!} \;>\; \sum_{\ell=1}^{r-1} \frac{\ell^N}{(\ell-1)!}$$

then there will exist a non-trivial linear combination of the congruences with a vanishing right hand side. $\qquad\square$

Lemma 4.6 allows us to express an atom $m(y_1) = m(a_1 b_1)$ as t/s, where t and s are polynomials in atoms in more variables, but of lower degrees in each of the variables. The next lemma keeps track of the way the number of variables grows.

LEMMA 4.7. — *Let p and q be integers. Assume $p > 0$, $q \geqslant 0$. Suppose we have a monomial*

$$y_1 = x_{10}^{\varepsilon_0} x_{11}^{\varepsilon_1} \cdots x_{1(p+q)}^{\varepsilon_{p+q}}$$

so that:

(i) *For $0 \leqslant j \leqslant p$ we have $0 \leqslant \varepsilon_j \leqslant \ell$.*

(ii) *For $p+1 \leqslant j \leqslant p+q$ we have $\varepsilon_j \in \{0,1\}$.*

Suppose further that we have variables $x_{10}, \ldots, x_{1(p+q+N)}$ with $N = N(r)$ as in Lemma 4.6. In other words, $n \geqslant p + q + N$. Then there exists an identity in $S_r^{\Sigma_r}$

$$s \cdot m(y_1) \;=\; t$$

with $s \neq 0$. Both s and t are polynomials in atoms of the form

$$m(x_{10}^{\delta_0} x_{11}^{\delta_1} \cdots x_{1(p+q+N)}^{\delta_{p+q+N}})$$

where

(i) *For $0 \leqslant j \leqslant p$ we have $0 \leqslant \delta_j \leqslant (\ell+1)/2$.*

(ii) *For $p+1 \leqslant j \leqslant p+q+N$ we have $\delta_j \in \{0,1\}$.*

(iii) *In the interval $p+1 \leqslant j \leqslant p+q$, the number of j with $\delta_j = 1$ is $\leqslant (q+1)/2$.*

(iv) *For the atoms that occur in the expansion of s, $\delta_0 = \delta_1 = \ldots = \delta_{p+q} = 0$.*

Proof. — We are given $y_1 = x_{10}^{\varepsilon_0} x_{11}^{\varepsilon_1} \cdots x_{1(p+q)}^{\varepsilon_{p+q}}$. For each j with $0 \leqslant j \leqslant p$ write $\varepsilon_j = \delta_j + \delta_j'$, with $\delta_j, \delta_j' \leqslant (\ell + 1)/2$. For j with $p + 1 \leqslant j \leqslant p + q$, put $\varepsilon_j = \delta_j + \delta_j'$ with $\delta_j = 0$ if j is odd, $\delta_j' = 0$ if j is even. Now define

$$a_1 = x_{10}^{\delta_0} x_{11}^{\delta_1} \cdots x_{1(p+q)}^{\delta_{p+q}}$$

$$b_1 = x_{10}^{\delta_0'} x_{11}^{\delta_1'} \cdots x_{1(p+q)}^{\delta_{p+q}'}$$

Then $m(y_1) = m(a_1 b_1)$, and our lemma follows from Lemma 4.6.   □

THEOREM 4.8. — *Let $N = N(r)$ be as in Lemma 4.6. If $n \geqslant r+3N+1$, then the quotient field of $S_r^{\Sigma_r}$ is generated by the special atoms of degree $\leqslant r + 2N + 2$. (see Definition 2.6). Moreover, when we express an element $s \in S_r^{\Sigma_r}$ as a/b, with a and b polynomials in special atoms, we can choose b to depend only on the variables x_{ij} with $j \geqslant 1$ (that is, not on x_{i0}).*

Proof. — By Proposition 3.8 we know that the ring $S_r^{\Sigma_r}$ is generated by atoms $m(y_1)$, with y_1 a polynomial of degree $\leqslant r$. Choose such an $m(y_1)$; we need to express it in the form a/b, with a and b polynomials in the special atoms.

Since y_1 is a polynomial of degree $\leqslant r$ in the variables x_{1j}, at most r variables occur, and the degree of y_1 in each variable is $\leqslant r$. Reordering the variables we may assume y_1 is a polynomial in $x_{10}, x_{11}, \ldots, x_{1r}$. Note that we have taken $r+1$ variables. This is because we want x_{10} to be in the first group of variables, which do not occur in the denominators. In other words, we leave x_{10} alone, and relabel the other variables so that y_1 is a polynomial only in the first $r + 1$ variables.

Now we apply Lemma 4.7. We put $p = r$, $q = q_1 = 0$. The lemma tells us that we can write $m(y_1)$ as a/b. In this expression, b is a polynomial in special atoms not involving the variables x_{ij} with $j \leqslant r$. And a is a polynomial in atoms $m(\tilde{y}_1)$. Now $\tilde{y}_1$ is a polynomial of degree $\leqslant (r + 1)/2$ in the variables $x_{10}, x_{11}, \ldots, x_{1r}$, and of degree $\leqslant 1$ in at most $N+(q_1+1)/2$ more variables. Put $q_2 = N + (q_1 + 1)/2$.

We can iterate this, expressing $m(\tilde{y}_1)$ as a'/b', and so on. In each iteration the permissible degrees in $x_{10}, x_{11}, \ldots, x_{1r}$ drop, but the number

of variables rises. Let the number of permissible variables at the ℓth iteration, other than $x_{10}, x_{11}, \ldots, x_{1r}$, be $\leqslant q_\ell$. We have a recurrence relation

$$q_{\ell+1} = N + (q_\ell + 1)/2$$

with $q_1 = 0$. This makes $q_2 = N + 1/2$, $q_3 = N + N/2 + 1/2 + 1/4$, and in general

$$q_\ell \leqslant \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots \right) + N \left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots \right)$$

$$= 1 + 2N.$$

In Lemma 4.7 the number n needs to be chosen at least $p + q_\ell + N$. The number of variables in the atoms that occur is bounded by $1 + p + (q_\ell + 1)/2 + N$. In our case $p = r$ and $q_\ell \leqslant 1 + 2N$. Hence it suffices to have $n \geqslant r + 3N + 1$, and the atoms will be in $\leqslant r + 2N + 2$ variables. $\square$

This finishes the proof that the special atoms generate. Next we want to deduce the same for the extra special atoms. Before we launch into the general argument, it might be helpful to illustrate it in the case $r = 2$.

Illustration 4.9. — Suppose $r = 2$, and suppose z_1 is a monomial in the variables $x_{15}, x_{16}, \ldots, x_{1n}$. The sum of four terms

$$m(x_{10}x_{11}x_{13}z_1)m(x_{10}x_{12}x_{14}) \quad + \quad m(x_{10}x_{11}x_{13})m(x_{10}x_{12}x_{14}z_1)$$

$$- \quad m(x_{10}x_{11}x_{14}z_1)m(x_{10}x_{12}x_{13}) \quad - \quad m(x_{10}x_{11}x_{14})m(x_{10}x_{12}x_{13}z_1)$$

is equal to $z_1 + z_2$ times the expression

$$m(x_{10}x_{11}x_{13})m(x_{10}x_{12}x_{14}) \quad - \quad m(x_{10}x_{11}x_{14})m(x_{10}x_{12}x_{13}).$$

This permits us to express $m(z_1) = z_1 + z_2$ as a quotient, and if $m(z_1)$ is special then the quotient involves only extra special atoms.

Notation 4.10. — We begin with $3r^2$ variables y_{ij}, x_{ij} and $\tilde{x}_{ij}$ with $1 \leqslant i, j \leqslant r$. Let $\alpha \in \Sigma_r$ be a permutation on the set $\{1, 2, \ldots, r\}$. If j is an integer with $1 \leqslant j \leqslant r$, then $\alpha(j)$ will denote the image of j under the permutation α. We wish to consider the substitution ϕ_α given by

$$\phi_\alpha(y_{ij}) = x_{ij}\tilde{x}_{i,\alpha(j)}.$$

The map ϕ_α extends to a ring homomorphism $k[y_{ij}] \longrightarrow k[x_{ij}, \tilde{x}_{ij}]$. By abuse of notation we also call this ring homomorphism ϕ_α. We define the linear map $A : k[y_{ij}] \longrightarrow k[x_{ij}, \tilde{x}_{ij}]$ by the formula

$$A(Y) = \sum_{\alpha \in \Sigma_r} \mathrm{sgn}(\alpha) \cdot \phi_\alpha(Y)$$

where $\mathrm{sgn}(\alpha)$ means the sign of the permutation α.

We are particularly interested in $A(Y)$ for Y a monomial of degree r in the y_{ij}. More particularly we care about Y of the form

$$Y = Y_f = \prod_{j=1}^{r} y_{f(j),j}$$

for some function $f : \{1, 2, \ldots, r\} \longrightarrow \{1, 2, \ldots, r\}$.

LEMMA 4.11. — Let $Y_f = \prod_{j=1}^{r} y_{f(j),j}$ be as in Notation 4.10. If $f : \{1, 2, \ldots, r\} \longrightarrow \{1, 2, \ldots, r\}$ is not injective, then $A(Y_f) = 0$. The elements $A(Y_f)$, where f runs over the permutations $f \in \Sigma_r$, are linearly independent over k.

Proof. — If f is not injective, then there is a pair of distinct numbers $\ell \neq \ell'$ with $f(\ell) = f(\ell')$. Let σ be the transposition which switches ℓ and ℓ'. If $j \neq \ell, \ell'$, then $\alpha\sigma(j) = \alpha(j)$. And $\alpha\sigma(\ell) = \alpha(\ell')$, $\alpha\sigma(\ell') = \alpha(\ell)$. We have

$$\phi_\alpha(Y_f) = \prod_{j=1}^{r} x_{f(j),j} \prod_{j=1}^{r} \tilde{x}_{f(j),\alpha(j)} \qquad \phi_{\alpha\sigma}(Y_f) = \prod_{j=1}^{r} x_{f(j),j} \prod_{j=1}^{r} \tilde{x}_{f(j),\alpha\sigma(j)}$$

and we conclude that $\phi_\alpha(Y_f) = \phi_{\alpha\sigma}(Y_f)$. The sign of $\alpha\sigma$ is opposite to the sign of α, and so in $A(Y_f)$ these terms cancel. This is true for every right coset of the group $\{1, \sigma\} \subset \Sigma_r$, and hence $A(Y_f) = 0$.

Now assume f is a permutation. Then the first product in the expression

$$\phi_\alpha(Y_f) = \prod_{j=1}^{r} x_{f(j),j} \prod_{j=1}^{r} \tilde{x}_{f(j),\alpha(j)}$$

enables us to determine f, and the second product determines for us α. In other words, the monomials $\phi_\alpha(Y_f)$ are all distinct, as α, f vary over Σ_r. We have $(r!)^2$ different monomials $\phi_\alpha(Y_f)$.

The polynomial $A(Y_f)$ is a non-trivial linear combination of the $r!$ distinct monomials $\phi_\alpha(Y_f)$, with f fixed and α varying. As we vary f, the polynomials $\{A(Y_f), f \in \Sigma_r\}$ are linear combinations of non-intersecting sets of monomials; there is no possible linear dependence.   □

Now we wish to define the polynomial in y_{ij} to which we will apply the linear map A. It will be defined as a polynomial over the field $K = k(x_{10}, x_{20}, \ldots, x_{r0}, z_1, z_2, \ldots, z_r)$. In other words, $Y = Y(z_i, y_{ij}, x_{i0})$ will be a polynomial in the $r(r+2)$ variables z_i, y_{ij}, x_{i0} with $0 \leqslant i, j \leqslant r$. As far as the linear map A is concerned that extra variables z_i and x_{i0} will be viewed as constants.

DEFINITION 4.12. — Let z_i, y_{ij}, x_{i0} be variables with $1 \leqslant i, j \leqslant r$. We study the following expression:

$$Y(z_i, y_{ij}, x_{i0}) = \sum_{j=1}^{r} m(x_{10}y_{11}) \cdots m(x_{10}y_{1j}z_1) \cdots m(x_{10}y_{1r}).$$

Illustration 4.13. — It might be helpful to treat the case $r = 2$, as in Illustration 4.9. In this case $Y = Y(z_i, y_{ij}, x_{i0})$ becomes

$$Y \quad = \quad m(x_{10}y_{11}z_1)m(x_{10}y_{12}) \quad + \quad m(x_{10}y_{11})m(x_{10}y_{12}z_1).$$

This makes $A(Y)$ equal to the expression

$$A(Y) = m(x_{10}x_{11}\tilde{x}_{11}z_1)m(x_{10}x_{12}\tilde{x}_{12}) + m(x_{10}x_{11}\tilde{x}_{11})m(x_{10}x_{12}\tilde{x}_{12}z_1)$$

$$- m(x_{10}x_{11}\tilde{x}_{12}z_1)m(x_{10}x_{12}\tilde{x}_{11}) - m(x_{10}x_{11}\tilde{x}_{12})m(x_{10}x_{12}\tilde{x}_{11}z_1),$$

and this agrees with the four–term sum in Illustration 4.9, up to the substitution

$$\tilde{x}_{i1} = x_{i3}, \qquad \tilde{x}_{i2} = x_{i4}.$$

LEMMA 4.14. — With $Y = Y(z_i, y_{ij}, x_{i0})$ as in Definition 4.12, we have

$$A(Y) = (z_1 + z_2 + \ldots + z_r) \cdot P(x_{ij}, \tilde{x}_{ij}, 1 \leqslant i, j \leqslant r) \cdot \prod_{i=1}^{r} x_{i0}$$

with $P(x_{ij}, \tilde{x}_{ij}, 1 \leqslant i, j \leqslant r)$ some non-zero polynomial.

Proof. — In $Y(z_i, y_{ij}, x_{i0})$, the coefficient of

$$x_{10}^{\varepsilon_1} x_{20}^{\varepsilon_2} \cdots x_{r0}^{\varepsilon_r}$$

is a sum of terms $z_\ell Y_f$, with $f : \{1, 2, \ldots, r\} \longrightarrow \{1, 2, \ldots, r\}$ so that there are ε_i distinct integers j with $f(j) = i$. By Lemma 4.11, $A(Y_f)$ vanishes unless f is a permutation. Hence only the term where $\varepsilon_i = 1$ for all i can have a non-zero contribution to $A(Y)$.

Therefore, consider the coefficient in $Y = Y(z_i, y_{ij}, x_{i0})$ of the monomial

$$x_{10} x_{20} \cdots x_{r0}.$$

It can be written as $(z_1 + z_2 + \ldots + z_r) \cdot Q(y_{ij})$, with $Q(y_{ij}) \neq 0$. Now $Q(y_{ij})$ is a non-zero linear combination of terms Y_f, with $f : \{1, 2, \ldots, r\} \longrightarrow \{1, 2, \ldots, r\}$ a permutation. Lemma 4.11 tells us that the $A(Y_f)$ are linearly independent, and hence $A\big(Q(y_{ij})\big)$ must be a non-zero polynomial in $x_{ij}, \tilde{x}_{ij}$, $1 \leqslant i, j \leqslant r$. We conclude that

$$A(Y) = (z_1 + z_2 + \ldots + z_r) \cdot \prod_{i=1}^{r} x_{i0} \cdot A\big(Q(y_{ij})\big),$$

as required.   □

THEOREM 4.15. — *Let $N = N(r)$ be as in Lemma 4.6. If $n \geqslant \max(r + 3N + 1, 3r + 2N + 2)$, then the quotient field of $S_r^{\Sigma_r}$ is generated by the extra special atoms (see Definition 2.6). Moreover, when we express an element $s \in S_r^{\Sigma_r}$ as a/b, with a and b polynomials in extra special atoms, we can choose b to be of the form*

$$b = \prod_{i=1}^{r} x_{i0} \cdot P(x_{ij}, 1 \leqslant j).$$

Proof. — Theorem 4.8 reduces us to showing that any special atom $m(z_1)$ of degree $\leqslant r + 2N + 2$ can be expressed in the form a/b above. If z_1 is divisible by x_{10} there is nothing to prove, since $m(z_1)$ is extra special. Reordering the variables x_{ij} with $j \geqslant 1$ if necessary, we may assume z_1 is a monomial in $x_{1\ell}$, with $2r + 1 \leqslant \ell \leqslant 3r + 2N + 2$. Note that n is large enough so we can do that.

Now we wish to apply Lemma L3.10. To do so, put $\tilde{x}_{ij} = x_{i(j+r)}$. We have a formula

$$A\big(Y(z_i, y_{ij}, x_{i0})\big) \quad = \quad (z_1 + z_2 + \ldots + z_r) \cdot P(x_{ij}, \tilde{x}_{ij}, 1 \leqslant i, j \leqslant r) \cdot \prod_{i=1}^{r} x_{i0}$$

and hence

$$\frac{A\big(Y(z_i, y_{ij}, x_{i0})\big)}{A\big(Y(1, y_{ij}, x_{i0})\big)} \quad = \quad \frac{z_1 + z_2 + \ldots + z_r}{r} \ .$$

That is, we have explicitly expressed $m(z_1) = z_1 + z_2 + \ldots + z_r$ as a quotient. Note that $A\big(Y(z_i, y_{ij}, x_{i0})\big)$ and $A\big(Y(1, y_{ij}, x_{i0})\big)$ are both sums of products of atoms $m(x_{10} x_{1j} \tilde{x}_{1\ell})$ and $m(x_{10} x_{1j} \tilde{x}_{1\ell} z_1)$, with $1 \leqslant j, \ell \leqslant r$. All of these are extra special. $\qquad\square$

BIBLIOGRAPHY

[1] M. V. CATALISANO, A. V. GERAMITA, A. GIMIGLIANO, Ranks of tensors, secant varieties of Segre varieties and fat points, Linear Algebra Appl. 355 (2002), 263–285.

[2] M. V. CATALISANO, A. V. GERAMITA, A. GIMIGLIANO, Erratum to: "Ranks of tensors, secant varieties of Segre varieties and fat points" [Linear Algebra Appl. 355 (2002), 263–285; MR 2003g:14070], Linear Algebra Appl. 367 (2003), 347–348.

[3] L. D. GARCIA, M. STILLMAN, B. STURMFELS, Algebraic geometry of Bayesian networks, (e-print http://arXiv.org/abs/math.AG/0301255).

[4] I. M. GEL′FAND, M. M. KAPRANOV, A. V. ZELEVINSKY, Discriminants, resultants, and multidimensional determinants, Mathematics: Theory & Applications, Birkhäuser Boston Inc., Boston, MA, 1994.

[5] L. A. GOODMAN, Exploratory latent structure analysis using both identifiable and unidentifiable models, Biometrika 61 (1974), 215–231.

[6] P. HALL, X.-H. ZHOU, Nonparametric estimation of component distributions in a multivariate mixture, Ann. Statist. 31 (2003), 201–224.

[7] P. HALL, A. NEEMAN, R. PAKYARI, R. ELMORE, Nonparametric inference in multivariate mixtures, To appear.

[8] J. M. LANDSBERG, L. MANIVEL, On the ideals of secant varieties to Segre varieties, (e-print http://arXiv.org/abs/math.AG/0311388).

[9] B.G. LINDSAY, "Mixture Models: Theory, Geometry, and Applications," Hayward, Institute of Mathematical Statistics, 1995.

[10] G.J. MAC LACHLAN, D. PEEL, "Finite Mixture Models," John Wiley & Sons, 2000.

[11] A. MATTUCK, The field of multisymmetric functions, Proc. Amer. Math. Soc. 19 (1968), 764–765.

[12] M. NAGATA, On the normality of the Chow variety of positive 0-cycles of degree m in an algebraic variety, Mem. Coll. Sci. Univ. Kyoto. Ser. A. Math. 29 (1955), 165–176.

[13] A. NEEMAN, Zero cycles in $\mathbb{P}^n$, Advances in Math. 89 (1991), 217–227.

[14] E. NETTO, Vorlesungen über Algebra, Teubner Verlag, Leipzig, 1896.

[15] H. TEICHER, Identifiability of mixtures, Ann. Math. Statist. 32 (1961), 244–248.

[16] H. TEICHER, Identifiability of finite mixtures, Ann. Math. Statist. 34 (1963), 1265–1269.

[17] D.M. TITTERINGTON, A.F. SMITH, U.E. MAKOV, "Statistical Analysis of Finite Mixture Distributions," John Wiley & Sons, 1985.

[18] H.WEYL, The Classical Groups. Their Invariants and Representations, Princeton University Press, Princeton, N.J., 1939.

[19] S.J. YAKOWITZ, J. D. SPRAGINS, On the identifiability of finite mixtures, Ann. Math. Statist. 39 (1968), 209–214.

Ryan ELMORE, Peter HALL, Amnon NEEMAN
The Australian National University
Mathematical Sciences Institute
Centre for Mathematics and its Applications
John Dedman Building
Canberra, ACT 0200 (Australia).

Ryan.Elmore@anu.edu.au
Peter.Hall@anu.edu.au
Amnon.Neeman@anu.edu.au