Secant Varieties and Statistical Models

Elizabeth S. Allman

University of Alaska Fairbanks

January 8, 2008

Joint Mathematics Meetings

Joint work with

J. Rhodes — Mathematics and Statistics

UNIVERSITY OF ALASKA FAIRBANKS

C. Matias — CNRS, Laboratoire

Statistique et Génome

Background:

"An application of classical invariant theory to identifiability in non-parametric mixtures"

R. Elmore, P. Hall, and A. Neeman, Annales de L'institut Fourier, (2005) $\mathbf{55}(1)$, 1–28.

secant varieties $\leftrightsquigarrow$ identifiability of statistical parameters

Example

For a population, 3 observed variables X, Y, Z.

1. height < 6 ft (Y$= 1$, N$= 2$)
2. systolic blood pressure < 120 (Y$= 1$, N$= 2$)
3. eye color (Br$= 1$, Bl$= 2$, Gr$= 3$)

In addition,

one hidden variable W, with $r = 2$ states: (ethnic group)

This can be formulated as a *conditional independence* model, with 9 parameters:

- sub-pop proportions for ethnic groups:

$$\pi_1 = \text{Prob}(\ W = 1\ ),$$
$$\pi_2 = 1 - \pi_1 = \text{Prob}(\ W = 2\ )$$

- probabilities of observations for sub-pop i:

$$\begin{cases} x_{ij} = \text{Prob}(X = j \mid W = i),\ j = 1, 2 \\ y_{ij} = \text{Prob}(Y = j \mid W = i),\ j = 1, 2 \\ z_{ij} = \text{Prob}(Z = j \mid W = i),\ j = 1, 2, 3 \end{cases}$$

Joint distribution is given by the 12 quantities:

$$p_{ijk} = \text{Proportion of population with } X = i,\ Y = j,\ Z = k$$

which the model predicts to be

$$p_{ijk} = \pi_1\, x_{1i}\, y_{1j}\, z_{1k} + \pi_2\, x_{2i}\, y_{2j}\, z_{2k}$$

Identifiability —

Is it possible to identify the parameters from these twelve quantities?

Identifiability is required for statistical consistency of inference methods such as Maximum likelihood.

A more geometric viewpoint:

In $\mathbb{P}^1$, $\mathbb{P}^2$, consider:

$$\begin{cases} \mathbf{x}_1 = (x_{11}\ x_{12}) \\ \mathbf{y}_1 = (y_{11}\ y_{12}) \\ \mathbf{z}_1 = (z_{11}\ z_{12}\ z_{13}) \end{cases} \qquad \begin{cases} \mathbf{x}_2 = (x_{21}\ x_{22}) \\ \mathbf{y}_2 = (y_{21}\ y_{22}) \\ \mathbf{z}_2 = (z_{21}\ z_{22}\ z_{23}) \end{cases}$$

The $2 \times 2 \times 3$ tensors

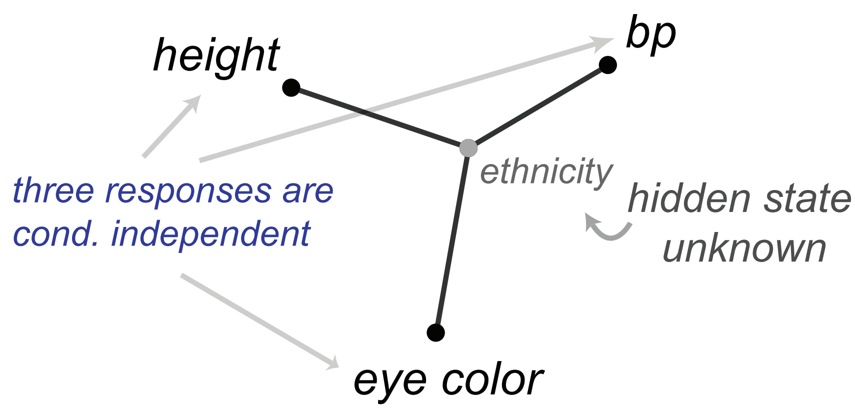
$$\mathbf{x}_1 \otimes \mathbf{y}_1 \otimes \mathbf{z}_1, \qquad \mathbf{x}_2 \otimes \mathbf{y}_2 \otimes \mathbf{z}_2$$

give the distributions for sub-pops 1, 2. Thus

$$\pi_1\, \mathbf{x}_1 \otimes \mathbf{y}_1 \otimes \mathbf{z}_1 + \pi_2\, \mathbf{x}_2 \otimes \mathbf{y}_2 \otimes \mathbf{z}_2$$

is the parameterized distribution for the population as a whole.

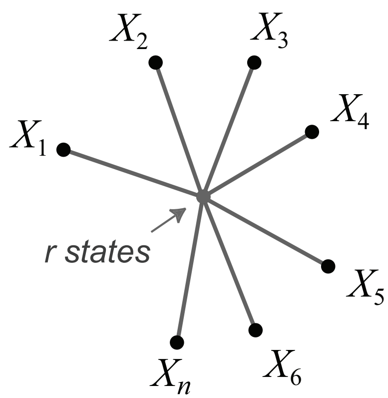
Tensor products describe independence, secants describe conditional independence



The distribution lies in

$$\mathrm{Sec}(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^2).$$

More generally:



- r state hidden variable $\rightsquigarrow$ higher secant
- n observed variables X_j, $j = 1, \ldots, n$, with s_j states

  $\rightsquigarrow$ n projective spaces $\mathbb{P}^{s_j - 1}$

$\mathcal{M}(r; s_1, \ldots, s_n)$ denotes the statistical model corresponding to $V = \mathrm{Sec}^r(\mathbb{P}^{s_1-1} \times \cdots \times \mathbb{P}^{s_n-1})$.

Note the particular parameterization is meaningful for statistics.

Questions:

- When is the dimension of V as *'expected'*?

  Much recent work, including

  Catalisano, Geramita, Gimigliano

  Abo, Ottaviani, Peterson

- What are generators for the ideals of V?

  Garcia, Stillman, Sturmfels

  Landsberg, Manivel, Weyman

  Allman, Rhodes

- Identifiability of parameters,

  J. Kruskal,

  J. Chang,

  Elmore, Hall, Neeman

Note: Understanding dimension $=$

understanding if parameterization is generically finite

This is *not* enough for statistical identifiability.

Label swapping for hidden variable $\leftrightsquigarrow$ at least a $r!$-to-one map

Want to understand

— structure of generic fiber

— characterize exceptional points, and their fibers

EHN thm — n binary observed variables:

$$V = \mathrm{Sec}^r(\underbrace{\mathbb{P}^1 \times \cdots \times \mathbb{P}^1}_{n})$$

Modify parameterization Φ', to account for $r!$-to-1-ness

(symmetrize parameter space)

Theorem (Elmore, Hall, Neeman): For each $r \geq 2$ 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.

Moreover,

$$\underbrace{c_1 \ln r}_{\text{parameter count}} \leq C(r) \leq \underbrace{c_2\, r \ln r}_{\text{EHN bound}}$$

Statistical interpretation:

$$\begin{array}{ccc} \text{Parameters} & \xrightarrow{\Phi'} & \mathrm{Sec}^r(\underbrace{\mathbb{P}^1 \times \cdots \times \mathbb{P}^1}_{n}) \\ \text{Stochastic Parameters} & \longrightarrow & \text{Joint Distributions} \end{array}$$

For r sub-populations, if the number n of observed binary variables is 'large,' then parameters are (generically) identifiable up to label swapping.

'Large' means $n > c_2\, r \ln r$.

(Bound here is poor.)

Desirable extensions:

- required number of observed variables should be reduced
- consider observed variables with more than 2 states:

  $$\mathcal{M}(r; s_1, \ldots, s_n) \longleftrightarrow \mathrm{Sec}^r(\mathbb{P}^{s_1-1} \times \cdots \times \mathbb{P}^{s_n-1})$$

- allow application to identifiability of many statistical models with hidden variables:
  - multivariate Bernoulli mixture distributions
  - hidden Markov models
  - random graph models, etc.

Basic tool:

Kruskal's Theorem (1977)

Consider the model $\mathcal{M}(r; s_1, s_2, s_3)$ with 3 observed variables.

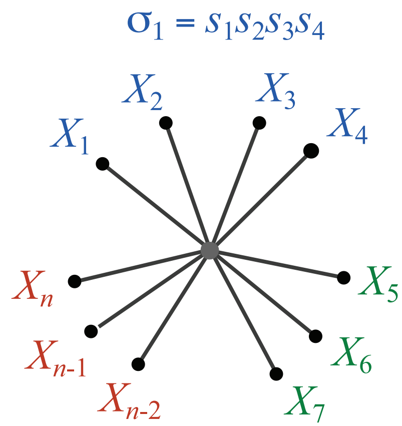
Let $\boldsymbol{\pi} = (\pi_1 \ \ldots \ \pi_r)$ be the sub-pop proportions, and M_j the $r \times s_j$ Markov matrix whose rows describe the jth observed variable for each sub-pop.

Define $I_j = \max\{k \mid$ *every* set of k rows of M_j is independent$\}$.

Theorem: If all entries of $\boldsymbol{\pi}$ are non-zero and

$$I_1 + I_2 + I_3 \geq 2r + 2,$$

then from the model's probability distribution $\boldsymbol{\pi},\ M_1,\ M_2,\ M_3$ are uniquely determined, up to some permutation.



Theorem (A–, Matias, Rhodes)

Consider the model $\mathcal{M}(r; s_1, s_2, \ldots, s_n)$ where $n \geq 3$. Suppose there exists a tripartition of the set $S = \{1, 2, 3, \ldots, n\}$ into three subsets S_1, S_2, S_3, such that if $\sigma_i = \prod_{j \in S_i} s_j$ then

$$\min(r, \sigma_1) + \min(r, \sigma_2) + \min(r, \sigma_3) \geq 2r + 2.$$

Then the model is generically identifiable, up to label swapping.

Cor. The r-class, n-binary-feature model $\mathcal{M}(r; 2, 2, \ldots, 2)$ is generically identifiable, up to label swapping, provided

$$n > 2\lceil \log_2 r \rceil,$$

i.e. $C(r) = \Theta(\ln r)$.

Cor. (in progress...) Applications to identifiability of hidden Markov models, random graph models etc.