Algebraic Causality

Eva Riccomagno

Department of Statistics
University of Warwick
Coventry CV4 7AL

emr@stats.warwick.ac.uk

Joint work with David Mond and Jim Smith

Main argument

Polynomial algebra is a natural framework for describing and manipulating (functional) relationships among a finite set of finite, discrete random variable $(X_1, \ldots, X_n)$.

Starting example: graphical models.

In particular the notion of causality (Pearl *Causality. Models, reasoning, and inference*, 2000) is polynomial in essence.

Thus algebraic geometry, the language of polynomials and varieties, should/could be employed.

References

Riccomagno and Smith *Non-graphical causality: a generalisation of the concept of a total cause*

Mond, Riccomagno and Smith *Algebraic causality*

Contents

Primitive probabilities

A factorisation of the joint mass function of $(X_1, \dots, X_n)$

$$p(X_1 = x_1, \dots, X_n = x_n) = p(x_1, \dots, x_n) =$$

$$p(X_1 = x_1) \prod_{i=2}^{n} p\left(X_i = x_i | X_{i-1} = x_{i-1}, \dots, X_1 = x_1\right)$$

for x_i taking a finite set of values $i = 1, \dots, n$

Power products in $p\left(x_i | \dots\right)$

Usually, the specific problem suggests (at least) a pre-ordering on $X_1, \dots, X_n$ that, in turn, suggests a factorisation and thus the primitive probabilities.

C.I. statements embedded by reducing the number of conditioning variables: pattern

---

In DAGS: $p(x_1, \dots, x_n) = \prod_{i=1}^{n} p(x_i | pa(x_i))$

(homogeneity of factorisation across settings)

Generalisation: different factorisation for different settings

Manifest variables

• For DAGs Pearl (1993,1995,2000) assumes to observe the joint mass function of a subset of m variables $\{X_{i_1}, \ldots, X_{i_m}\}$

$$p(x_{i_1}, \ldots, x_{i_m}) = \sum_{x_j \neq x_{i_1}, \ldots, x_{i_m}}^{n} \prod_{j=1}^{n} p(x_j | pa(x_j))$$

• Similarly we assume to observe some functional relationships among the primitive probabilities

$$m_i = f_i(p(X_1 = x_1, \ldots, X_n = x_n))$$

f_i polynomials.

P.S. A function on a finite set of points is representable as a polynomial.

Atomic intervention

Let $X_j \in \{X_1, \ldots, X_n\}$.

• For DAGs modify the joint mass function

$$p(x_1, \ldots, x_n || x_j = x'_j) = \prod_{i=1, i \neq j}^{n} p(x_i | pa'(x_i))$$

where $pa'(x_i)$ is obtained from $pa(x_i)$ replacing occurrences of x_j by x'_j

• Similarly, define w.r.t. the causal partial ordering the post-intervention distribution

$$\begin{aligned}\mathbf{X}_{\bar{a}} &= \{X_i \in \{X_1, \ldots, X_n\} : X_i \succ X_j\} \\ \mathbf{X}_a &= \{X_1, \ldots, X_n\} \setminus (\{X_j\} \cup \mathbf{X}_{\bar{a}})\end{aligned}$$

then

$$\begin{aligned}&p(X_1 = x_1, \ldots, X_n = x_n || X_j = x'_j) \\ &= p(\mathbf{x}_{\bar{a}}) p(\mathbf{x}_a | X_j = x'_j, \mathbf{x}_{\bar{a}})\end{aligned}$$

Chose the causal partial ordering compatible with the factorisation ordering

Example

X, Y, Z binary random variables

$$p(x, y, z) = p(x)p(y|x)p(z|x, y)$$

$$\begin{tikzcd} \mathbb{C}^7 \arrow[r] \arrow[d, leftrightarrow] & \mathbb{C}^3 \arrow[d, leftrightarrow] \\ {}^7 \arrow[r] & {}^3 \end{tikzcd}$$

Do $X = 1$ as projection in the hypercube over the face $p(x) = 1$

$$C^7 \cap \{p(x) = 1\} = C^3$$

Conditioning on $X = 1$ simplicial projection

$$p(y, z||X = 1) = \frac{p(X = 1, y, z)}{\sum_{y,z} p(X = 1, y, z)}$$

Cause-effect

The effect of the intervention $X_j = x'_j$ on a vector $Y \subset \{X_1, \ldots, X_n\} \setminus \{X_j\}$ is defined by

$$p(y||x'_j) = \sum_{'x_i \notin Y'} p(x_1, \ldots, x_n||x'_j)$$

Algebraic identifiability

The total cause C is *algebraically identified* by the observations M if and only if C can be written as a function of the observations M.

Pearl's problem

- Given full probability tables on a *subset* of variables.

- Interested in the "total cause" of X_j on X when we manipulate X_j to each possible value.

Algebraic Causality

- Pearl's concept of identifiability of a total cause is fundamentally an algebraic one (with links to elimination theory, change of variables, G-invariant polynomials when G is a finite group, ...)

- It is *only* graphical in the sense that Pearl uses the partial ordering of the graph to define his manipulation effects.

Applications

- parametric discrete Bayesian networks, parametric family assumptions
- censoring of sample information/non symmetrical information
- causal factorisation applying only in certain settings/partial causal graphs
- factorisation not associated with Bayesian networks, e.g. chain graph models
- causality associated with a limited class of actions/limited manipulation
- binary directed nests with hidden variables
- ...

An algorithm for identifiability of a total cause

Elimination theory from algebraic geometry

- $\mathbb{R}[z_1, \ldots, z_d]$ where z_i represent primitive probabilities
- Consider the factorisation of the joint into power products

$$p(x_1, \ldots, x_n) = \sum_{i=1}^{n} p(x_i | \ldots)$$

  possibly different across settings

- the polynomial giving the sum to one condition

- ...

- Define the extra indeterminates $m_1, \ldots, m_l$ giving the polynomials that define the manifest variables
- Consider $f = f(z_1, \ldots, z_d)$

- f is algebraically identified iff there exists $\bar{f}$ such that $f = \bar{f}(m_1, \ldots, m_l)$

in theory

in practice ...

Conclusion

1. Represent the class of models associated with these factorisations as a set of polynomials whose arguments are the primitive probabilities.

2. Identify the manifest variables as polynomial functions of the primitive probabilities.

3. Define the do-polynomials by lifting the definition valid for graphical B.N.'s and by transforming the factorisation of the joint mass function of $\{X_1, \ldots, X_n\}$ to an analogous post-intervention factorization.

4. Determine whether the effect of a total cause is identifiable by investigating algebraically whether the do-polynomials can be written explicitly as functions of manifest variables.

5. Exploit the algebraic symmetries inherent in the functions defining the factorisation, the post-intervention factorisation and the manifest variables together with general theorems to simplify this task.

6. To address the residual geometry, inherent to a given problem, employ methods from computational algebraic geometry to determine identifiability.

In progress

- Non graphical models
- Intervention on a hidden node
- Symmetries/group invariance
- Implementation
- Marginalisation of hidden nodes