Staged Trees and Equivalence Classes of Chain Event Graph Models

Christiane Görgen and Jim Q. Smith

Department of Statistics, University of Warwick, Coventry CV4 7AL, UK. `c.gorgen@warwick.ac.uk, j.q.smith@warwick.ac.uk`

ABSTRACT We are interested in characterising the class $[\mathcal{T},\Theta_\mathcal{T}]$ of *statistically equivalent* representations of a probability tree model $\mathbb{P}_{(\mathcal{T},\Theta_\mathcal{T})}$. This problem is analogous to the one of finding one *essential graph* $D^\star$ which indexes a class $[D]$ of DAG representations for a BN model $\mathbb{P}_D$. In the case of staged trees (and chain event graphs), no such graphical result is available. We thus present an alternative algebraic characterisation, showing that an *interpolating polynomial* $c_\mathcal{T} \in \mathbb{R}[\Theta_\mathcal{T}]$ uniquely identifies $[\mathcal{T},\Theta_\mathcal{T}]$. This is important for computational reasons, in model selection and for causal discovery.

The Probability Tree

Let the graph $\mathcal{T} = (V,E)$ be an *event tree* and $\Theta_\mathcal{T} = \{\theta_v \mid v \in V\}$ a set of parameter vectors $\theta_v = (\theta(e) \mid e \in E(v)) \in \Delta^\circ_{\#E(v)-1}$. Then,

$$\pi_\theta(\lambda) = \prod_{e\in E(\lambda)} \theta(e)$$

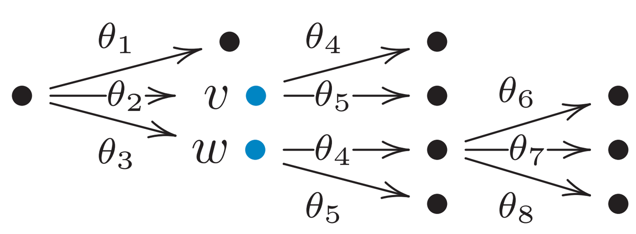
defines a strictly positive probability mass function on the root-to-leaf paths $\lambda \in \Lambda(\mathcal{T})$. We call the pair $(\mathcal{T},\Theta_\mathcal{T})$ a *probability tree*.

Such a labelled graph is a picture for a discrete parametric statistical model

$$\mathbb{P}_{(\mathcal{T},\Theta_\mathcal{T})} = \left\{ \underset{\sim}{\pi_\theta} \mid \theta \in \Delta^\circ_{d(E)} \right\} \subseteq \Delta^\circ_{\#\Lambda(\mathcal{T})-1}.$$

Denote by $[\mathcal{T},\Theta_\mathcal{T}]$ the set of possible representations of $\mathbb{P}_{(\mathcal{T},\Theta_\mathcal{T})}$. We say that two trees $(\mathcal{T},\Theta_\mathcal{T})$, $(\mathcal{S},\Theta_\mathcal{S}) \in [\mathcal{T},\Theta_\mathcal{T}]$ are *statistically equivalent*.

Two vertices $v, w \in V$ are in the same *stage* if and only if their parameter vectors coincide, $\theta_v = \theta_w$. We then write $v \sim w$.



Think of this as a sort of *conditional independence* statement, as formalised in Thwaites & Smith (2015).

The stage structure of a tree $(\mathcal{T},\Theta_\mathcal{T})$ may be captured in the *stage ideal*

$$I_\mathcal{T} = \sum_{v\sim w} \langle \theta_{v,i} - \theta_{w,i} \mid i = 1,\ldots,\#E(v_i)\rangle$$

in a ring $\mathbb{R}[\Theta_\mathcal{T}]$. This object is geometrically very simple but not invariant across $[\mathcal{T},\Theta_\mathcal{T}]$.

- Finite, discrete BN models form a *subclass* of the class of staged probability trees.
- Staged trees are graphically more complex but much more expressive than BNs.
- These trees have the same algebraic properties as *Chain Event Graphs*, introduced in Smith & Anderson (2008).

Polynomials in Tree Models

We define an *interpolating polynomial* of the model $\mathbb{P}_{(\mathcal{T},\Theta_\mathcal{T})}$ as $c_{g,\mathcal{T}}(\theta) = \sum_{\lambda\in\Lambda(\mathcal{T})} g(\lambda)\pi_\theta(\lambda)$, where $g = 1$ or an indicator function. Denote by $[\mathcal{T},\Theta_\mathcal{T}]^c \subseteq [\mathcal{T},\Theta_\mathcal{T}]$ the class of *polynomially equivalent* tree representations, sharing the same interpolating polynomial.

There is a bijective map $\mathfrak{c} : s(c(\theta)) \mapsto (\mathcal{T},\Theta_\mathcal{T})$, identifying a tree representation from a *tree-compatible* polynomial c. Interpolating polynomials are tree-compatible. In practice, $\mathfrak{c}$ maps a certain order of summation to a corresponding graph, using $\theta(e) \mapsto e \in E$ (see below).

Main Results

PROPOSITION For a finite and discrete BN model with decomposable DAG D and class of tree representations $[\mathcal{T},\Theta_\mathcal{T}]_D$ we find that $[\mathcal{T},\Theta_\mathcal{T}]^c = [\mathcal{T},\Theta_\mathcal{T}]_D$ for any *clique-induced* interpolating polynomial c. As a consequence, one interpolating polynomial captures all tree representations of a decomposable BN and thus, in particular, all fatorisations according to D.

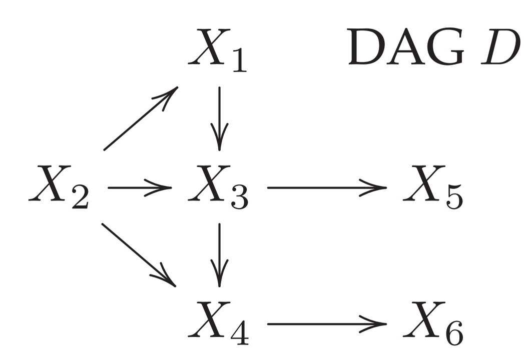
THEOREM Let $(\mathcal{S},\Theta_\mathcal{S}), (\mathcal{T},\Theta_\mathcal{T}) \in [\mathcal{T},\Theta_\mathcal{T}]$ be statistically equivalent staged trees with interpolating polynomials $c_\mathcal{S} \in \mathbb{R}[\Theta_\mathcal{S}]$ and $c_\mathcal{T} \in \mathbb{R}[\Theta_\mathcal{T}]$. Then there is a bijective map $\Phi : c_\mathcal{S} \mapsto c_\mathcal{T}$. In particular, we can transform $(\mathcal{S},\Theta_\mathcal{S})$ into $(\mathcal{T},\Theta_\mathcal{T})$ using a finite number of tree-compatible reorderings of the interpolating polynomial in combinations with substitutions of terms of edge probabilities.

Example: Tree representations of a decomposable BN

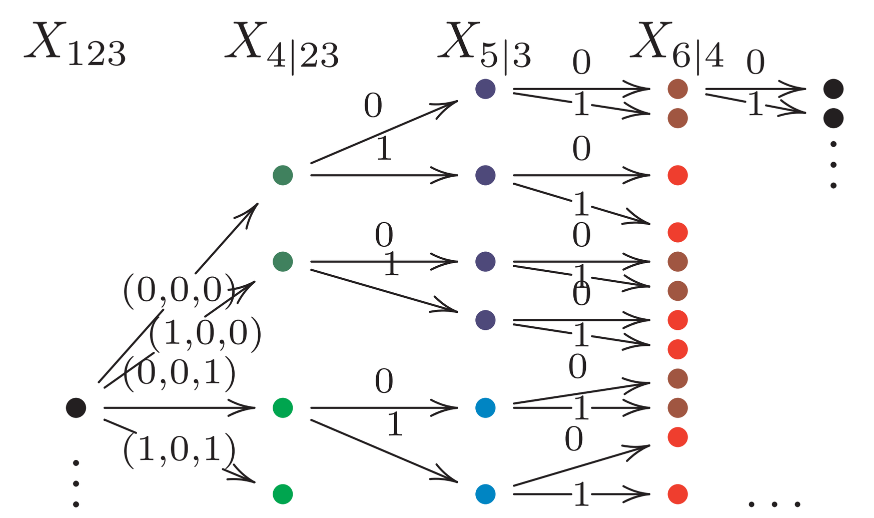
Consider a BN with (clique) parametrisation

$$p_\theta(x) = \theta(x_1,x_2,x_3)\theta(x_2,x_3,x_4)\theta(x_3,x_5)\theta(x_4,x_6)$$

according to a decomposable DAG D, shown on the right. Define a probability tree $(\mathcal{T},\Theta_\mathcal{T}) \in [\mathcal{T},\Theta_\mathcal{T}]_D$ with $\pi_\theta(\lambda(x)) = p_\theta(x)$ for all $x \in \mathbb{X}$.



The interpolating polynomial $c_\mathcal{T}(\theta) = \sum_{x\in\mathbb{X}} p_\theta(x)$ is given below. Each of the summations we see corresponds (via the map $\mathfrak{c}$ above) to a different tree representation in the class $[\mathcal{T},\Theta_\mathcal{T}]^c = [\mathcal{T},\Theta_\mathcal{T}]_D$.The interpolating polynomial allows us to traverse the whole class of tree representations of the BN, and *stratified* trees are the ones giving a recursive factorisation of p_θ according to D.



$$\begin{aligned}
c_\mathcal{T}(\theta) &= \sum_{(x_1,x_2,x_3)\in\mathbb{X}_{\{1,2,3\}}} \theta(x_1,x_2,x_3) \sum_{x_4\in\mathbb{X}_4} \theta(x_2,x_3,x_4) \sum_{x_5\in\mathbb{X}_5} \theta(x_3,x_5) \sum_{x_6\in\mathbb{X}_6} \theta(x_4,x_6) && p_{123}(x_1,x_2,x_3)p_4(x_4|x_2,x_3)p_5(x_5|x_3)p_6(x_6|x_4)\\
&= \sum_{(x_2,x_3,x_4)\in\mathbb{X}_{\{2,3,4\}}} \theta(x_2,x_3,x_4) \sum_{x_1\in\mathbb{X}_1} \theta(x_1,x_2,x_3) \sum_{x_5\in\mathbb{X}_5} \theta(x_3,x_5) \sum_{x_6\in\mathbb{X}_6} \theta(x_4,x_6) && p_{234}(x_2,x_3,x_4)p_1(x_1|x_2,x_3)p_5(x_5|x_3)p_6(x_6|x_4)\\
&= \sum_{(x_3,x_5)\in\mathbb{X}_{\{3,5\}}} \theta(x_3,x_5) \sum_{(x_1,x_2)\in\mathbb{X}_{\{1,2\}}} \theta(x_1,x_2,x_3) \sum_{x_4\in\mathbb{X}_4} \theta(x_2,x_3,x_4) \sum_{x_6\in\mathbb{X}_6} \theta(x_4,x_6) && p_{35}(x_3,x_5)p_{12}(x_1,x_2|x_3)p_4(x_4|x_2,x_3)p_6(x_6|x_4)\\
&= \sum_{(x_4,x_6)\in\mathbb{X}_{\{4,6\}}} \theta(x_4,x_6) \sum_{(x_2,x_3)\in\mathbb{X}_{\{2,3\}}} \theta(x_2,x_3,x_4) \sum_{x_1\in\mathbb{X}_1} \theta(x_1,x_2,x_3) \sum_{x_5\in\mathbb{X}_5} \theta(x_3,x_5). && p_{46}(x_4,x_6)p_{23}(x_2,x_3|x_4)p_{12}(x_1|x_2,x_3)p_5(x_5|x_3)
\end{aligned}$$

Conclusions

Our paper analyses discrete and context-specific BNs as well as stratified CEGs as subclasses of staged tree models, and characterises their equivalence classes via interpolating polynomials. We developed an algorithm for the map $\mathfrak{c}$, identifying corresponding graphs from a tree-compatible polynomial, and have examined computational aspects of our work in Leonelli et al. (2015). Future research will focus on causal interpretations of these results, based on Cowell & Smith (2015).

REFERENCES Christiane Görgen and Jim Q. Smith, *Equivalence Classes of Chain Event Graph Models*, in preparation.

Christiane Görgen, Manuele Leonelli and Jim Q. Smith, *A Differential Approach for Staged Trees*, Conference Proceedings of ECSQARU 2015.

Jim Q. Smith and Paul E. Anderson, *Conditional independence and Chain Event Graphs*, Journal of Artifical Intelligence 172, 42-68, 2008.

Robert G. Cowell and Jim Q. Smith, *Causal discovery through MAP selection of stratified Chain Event Graphs*, Electronic Journal of Statistics 8, 965-997, 2014.

Peter A. Thwaites and Jim Q. Smith, *A Separation Theorem for Chain Event Graphs*, submitted to the Electronic Journal of Statistics, 2015.

ALGEBRAIC STATISTICS 2015, June 8–11, Dipartimento di Matematica, Università degli Studi di Genova, Italia.

This research (with grant number EP/L505110/1) is funded by EPSRC Engineering and Physical Sciences Research Council.