The algebra of integrated partial belief systems

Manuele Leonelli, James Q. Smith

Department of Statistics, The University of Warwick, CV47AL Coventry, UK

Eva Riccomagno

Dipartimento di Matematica, Universita' degli Studi di Genova, Via Dodecaneso 35, 16146 Genova, Italia

Abstract

Current decision support systems address domains that are heterogeneous in nature and becoming larger. Such systems often require the input of expert judgement about a variety of different fields and an intensive computational power to produce scores to rank the available policies. The technology of integrating decision support systems has been recently extended to enable a formal distributed multi-agent decision analysis. Inference in these system is designed to be distributed so that for the sole purpose of decision support each panel needs to deliver only certain summaries of the variables under its jurisdiction. By using an algebraic approach, we are able to identify the required summaries and to demonstrate that coherence, in a sense we formalize here, is still guaranteed when panels only share a partial specification of their model with other panel members. We illustrate such algorithms for a variety of frameworks, including a specific class of Bayesian networks. For this class we derive closed form formulae for the computations of the joint moments of variables that determine the score of different policies.

Keywords: Bayesian networks, Integrating decision support systems, Polynomial algebra, Structural equation models

1. Introduction

Probabilistic decision support tools for single agents are now, although still being refined, well developed and used in practice in a variety of domains [14, 9]. However, the size of current applications often requires that expert judgements are delivered by diverse panels of experts each with their own particular domain knowledge. For instance, in nuclear emergency management judgements concerning issues such as the safety of the source term, the atmospheric dispersion of a cloud of contamination and the effects on human health deriving from radioactive intake, all need to be taken into account in the decision making process [11, 18].

Recently, integrating decision support systems (IDSSs) [12, 23] have been defined to extend coherence requirements traditionally applied within a Baye-sian decision support system for single agents so that it applies to this new multi-expert setting. To be practical such coherent systems need to be *distributed* in the sense that the overall scoring of the available policies can be uniquely deduced by the beliefs individually delivered by the panels. Under conditions formally and extensively discussed in [12] and [23], it is shown that a variety of both dynamic and non-dynamic graphical models can be used as an overarching integrating tool to provide a unique coherent picture of the whole problem, in such a way that the judgements of the different panels do not contradict each other. A variety of different methodologies can be employed by the panels to model the domain under their jurisdiction, as for example large scale hierarchical Bayesian spatio-temporal models based on advanced computational algorithms [1] or probabilistic emulators over massive deterministic simulators [6, 8].

In [23] we mostly focused on the inferential full-distributional difficulties associated to this integration. However, a formal Bayesian decision analysis is based on the maximization of an expected utility (EU) function that often only depends on some simple summaries of key output variables, for example a few

Preprint submitted to Elsevier *July 3, 2015*

low order moments. By requesting from the relevant panels only the value of these expectations, the implementation of an IDSS can become orders of magnitude more manageable. Panels then just need to communicate a few summaries of their analysis: a trivial and fast task to perform within most inferential systems.

Perhaps surprisingly, it is common to be able to define a coherent and distributed system with this property by only specifying qualitative relationships between its random variables and quantifying a few of their associated summaries. The required EU values can then often be calculated using familiar tower rules. The prospect of being able to build feasible and coherent decision support over huge systems is therefore now on the horizon.

The expected utilities of such an IDSS are usually polynomials whose indeterminates are functions of the panels' delivered summaries. This polynomial structure enables us to identify new separation conditions, often implicit in standard conditional independence over the parameters of certain graphical models [5, 24], that guarantee coherence in these types of distributed systems. Under these conditions, we develop new propagation algorithms for BNs, here called *algebraic substitutions*, for the distributed computations of an IDSS EU scores. This generalizes the theory of the computation of moments of decomposable functions [3, 17] to multilinear ones. The process of algebraic substitutions mirrors the recursions of [10] for the computation of the first two moments of chain graph models. Here, focusing only on certain BN models, we are able to explicitly compute any joint moment and provide an intuitive graphical interpretation of the propagation.

The paper is structured as follows. Section 2 reviews the main constituents of an IDSS and introduces an algebraic definition of expected utilities. Section 3 defines new separation conditions tailored to the needs of an IDSS. In Section 4 we prove that coherence can be retained in integrating systems under these milder conditions. Section 5 studies BNs and introduces algebraic substitutions. We conclude the paper with a discussion.

2. An algebraic description of integrating systems

Consider a random vector $\boldsymbol{Y} = (\boldsymbol{Y}_i^{\mathrm{T}})_{i\in[m]}$, $[m] = \{1, \dots, m\}$, where a subvector $\boldsymbol{Y}_i$ of $\boldsymbol{Y}$ is under the jurisdiction of a panel of experts G_i, $i \in [m]$. Let $\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}$ and $\boldsymbol{y}_i \in \boldsymbol{\mathcal{Y}}_i$ be instantiations of $\boldsymbol{Y}$ and $\boldsymbol{Y}_i$, respectively. Assume each panel of experts delivers beliefs about $\boldsymbol{\theta}_i$, the parameter of the density f_i over $\boldsymbol{Y}_i \mid (\boldsymbol{\theta}_i, d)$, where $d \in \mathcal{D}$ is one of the available policies in the decision space $\mathcal{D}$. Suppose $\boldsymbol{\theta}_i$ takes values in $\boldsymbol{\Theta}_i$ and let $\boldsymbol{\theta} = (\boldsymbol{\theta}_i^{\mathrm{T}})_{i\in[m]}$ take values in $\boldsymbol{\Theta}$. Let f, π_i and π denote densities over $\boldsymbol{Y} \mid (\boldsymbol{\theta}, d)$, $\boldsymbol{\theta}_i \mid d$ and $\boldsymbol{\theta} \mid d$, respectively. The implicit (although virtual) owner of the beliefs delivered by the panels will be referred to as the *supraBayesian* (SB).

The SB will process the panels' judgements in order to calculate various statistics of her reward vector $\boldsymbol{R}(\boldsymbol{Y}, d)$. Let $\boldsymbol{r} = \boldsymbol{r}(\boldsymbol{y}, d)$ denote an instantiation of $\boldsymbol{R}(\boldsymbol{Y}, d)$. For the purpose of a formal Bayesian analysis the SB will compute the set of *EU scores* $\{\bar{u}(d) : d \in \mathcal{D}\}$ as a function of both the utility function $u(\boldsymbol{r}, d)$ and the probability statements of the individual panels. The user will then be recommended to follow the policy d^* with the highest EU score, $\bar{u}(d^*)$, where the EU is computed as

$$\bar{u}(d) = \int_{\boldsymbol{\Theta}} \bar{u}(d \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid d)\mathrm{d}\boldsymbol{\theta}, \tag{1}$$

and

$$\bar{u}(d \mid \boldsymbol{\theta}) = \int_{\boldsymbol{\mathcal{Y}}} u(\boldsymbol{r}, d) f(\boldsymbol{y} \mid \boldsymbol{\theta}, d)\mathrm{d}\boldsymbol{y}, \tag{2}$$

is the *conditional expected utility (CEU)*. For simplicity and with no loss of generality we assume in this paper that $\boldsymbol{R} = \boldsymbol{Y}$.

2.1. How an integrating decision support system works

We first briefly review the IDSS theory (see [12, 23] for more details) relevant to this paper. An IDSS is defined by a set of agreements between the constituent panels concerning the qualitative structure of the decision problem addressed. Specifically they need to jointly agree on the available policies in the decision space $\mathcal{D}$, on the family of utility functions $\mathcal{U}$ supported by the system and on the dependence structure between various functions of $\boldsymbol{Y}$, $\boldsymbol{\theta}$ and d. This last agreement might be expressible, as we will see below, through a statistical graphical model for not only the distribution of $\boldsymbol{Y} \mid (\boldsymbol{\theta}, d)$, but also for the one over $\boldsymbol{\theta} \mid d$. The union of these three agreements is called **common-knowledge class (CK-class)**. Here everyone agrees on the components of the system and their relationships with each other. The CK-class defines the *qualitative* structure of the domain investigated and therefore more easily provides the framework for the group's agreement [21].

Given that the qualitative structure has been agreed within the CK-class, the *quantitative* belief specification in IDSSs is delegated to the most informed panel of experts about each given domain. These panels then individually deliver to the SB the necessary quantities for the computation of expected utilities only concerning the variables $\boldsymbol{Y}_i$ under their jurisdiction.

Although the IDSS is now fully defined by the CK-class together with the quantitative panels' specifications, there is no guarantee that the system is actually operationally useful. An IDSS will in general need to entertain the following condition.

Definition 1. An IDSS is said to be **adequate** for a CK-class if the SB can unambiguously calculate $\bar{u}(d)$ for any decision $d \in \mathcal{D}$ and any utility function $u \in \mathcal{U}$ from the beliefs of panel G_i, $i \in [m]$.

Without this property an IDSS would not be able to compute EU scores and produce a ranking of the various policies: therefore it would not be of any help to potential DMs. In [23] we introduce conditions that guarantee adequacy in a variety of inferential domains. In this paper, where we focus on decision making only, we introduce new conditions, often milder than those of [23], sufficient to guarantee that an IDSS is adequate for the task it was built.

2.2. Algebraic expected utilities and score separability

By approaching the theory of IDSSs from an algebraic viewpoint, we are able to identify the necessary panels' summaries and the required assumptions for adequacy. In order to do this we first need to define the EU polynomials.

Definition 2. The CEU $\bar{u}(d \mid \boldsymbol{\theta})$ of an IDSS is called *algebraic in the panels* if, for each $d \in \mathcal{D}$ and each panel G_i in charge of $\boldsymbol{Y}_i$ with parameter $\boldsymbol{\theta}_i$, $i \in [m]$, there exist $\lambda_i(\boldsymbol{\theta}_i, d)$ functions of $\boldsymbol{\theta}_i$ and d such that $\bar{u}(d \mid \boldsymbol{\theta})$ is a square-free polynomial q_d of the $\boldsymbol{\lambda}_i$

$$\bar{u}(d \mid \boldsymbol{\theta}) = q_d\left(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \cdots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d)\right).$$

Each $\boldsymbol{\lambda}_i$ is a vector of length s_i, where s_i is the number of summaries each panel is required to deliver. Let $\boldsymbol{\lambda}_i(\boldsymbol{\theta}_i, d) = (\lambda_{ji}(\boldsymbol{\theta}_i, d))_{j \in [s_i]}$, $[s_i]^0 = [s_i] \cup \{0\}$ and $\boldsymbol{b} \in B = \times_{i \in [m]} [s_i]^0$. For a given $\boldsymbol{b} = (b_i)_{i \in [m]}$ and $j \in [s_i]$, define $b_{j,i} = 0$ if $j \neq b_i$, $b_{j,i} = 1$ if $j = b_i$ and $b_{0,i} = 1$, for $i \in [m]$. Therefore $b_{j,i}$ is not zero if and only if either $j = 0$ or j equals the i-th entry of $\boldsymbol{b}$. Let $\lambda_{0i}(\boldsymbol{\theta}_i, d) = 1$, for every $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i$, $d \in \mathcal{D}$ and $i \in [m]$.

Example 1. Let $m = 2$, $s_1 = s_2 = 1$ and $\boldsymbol{b} = (0, 1)^{\mathrm{T}}$. Then $b_{0,1} = 1$, $b_{1,1} = 0$, $b_{0,2} = 0$ and $b_{1,2} = 1$.

Definition 3. The CEU $\bar{u}(d \mid \boldsymbol{\theta})$ of an IDSS is called *algebraic in the summaries*, or **algebraic**, if, for each $d \in \mathcal{D}$, q_d is a square-free polynomial of the λ_{ji}, $i \in [m]$, $j \in [s_i]^0$, such that

$$q_d\left(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d)\right) = \sum_{\boldsymbol{b} \in B} k_{\boldsymbol{b},d} \lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, d), \tag{3}$$

with $k_{\boldsymbol{b},d} \in \mathbb{R}$ and

$$\lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, d) = \prod_{i \in [m]} \prod_{j \in [s_i]^0} \lambda_{ji}(\boldsymbol{\theta}_i, d)^{b_{j,i}}.$$

Thus, $\lambda_{\boldsymbol{b}}$ is a monomial having at most one term not unity delivered by each panel and $k_{\boldsymbol{b},d}$ is a weight. For a given $\boldsymbol{b} \in B$, let

$$\mu_{ji}(d) = \mathbb{E}\left(\lambda_{ji}(\boldsymbol{\theta}_i, d)^{b_{j,i}}\right).$$

For the distributivity of the IDSS we need the following property.

Definition 4. Call an IDSS **score separable** if, in the notation above, all experts and the SB agree that, for all decisions $d \in \mathcal{D}$ and all indices $\boldsymbol{b} \in B$ such that $k_{\boldsymbol{b},d} \neq 0$,

$$\mathbb{E}\left(\lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, d)\right) = \prod_{i\in[m]} \prod_{j\in[s_i]^0} \mu_{ji}(d). \tag{4}$$

Let, for every $d \in \mathcal{D}$, $\boldsymbol{\mu}_i(d) = (\mu_{ji}(d))_{j\in[s_i]}$. A consequence of the definitions above is the following.

Lemma 1. *Suppose panel G_i delivers its vectors of expectations $\boldsymbol{\mu}_i(d)$, $i \in [m]$, $d \in \mathcal{D}$, to the SB. Then, assuming a CEU is algebraic, if the IDSS is score separable then it is adequate.*

PROOF. This follows from the definition of algebraic CEU in equation (3) and the definition of score separability.

We can therefore deduce from Lemma 1 that adequacy is guaranteed whenever score separability holds, under the assumption of an algebraic CEU. In the following section we introduce conditions that ensure this type of separability. We then identify classes of models that give rise to algebraic conditional expected utilities.

3. Moment and quasil independence

Equation (4) together with Lemma 1 shows that adequacy is guaranteed whenever the expectation of certain functions of the panels' parameters separate appropriately. We introduce now a new type of independence called quasi independence.

Definition 5. Let $q_d(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d))$ be the algebraic CEU of an IDSS. An IDSS is called **quasi independent** if

$$\mathbb{E}(q_d(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d))) = q_d(\mathbb{E}(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d)), \ldots, \mathbb{E}(\boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d))).$$

This condition requires the expectation of the product of certain functions of the parameters overseen by different panels to be equal to the product of the individual expectations.

Often the λ_{ji}, $i \in [m]$, $j \in [s_i]$, are monomial functions of the panels' parameters. It is therefore helpful to introduce the following independence condition specific for monomial functions. Let $<_{lex}$ denote a lexicographic order [4].

Definition 6. Let $\boldsymbol{\theta} = (\theta_i)_{i\in[n]} \in \mathbb{R}^n$ be a parameter vector and $\boldsymbol{c} = (c_i)_{i\in[n]} \in \mathbb{Z}^n_{\geq 0}$. We say that $\boldsymbol{\theta}$ entertains **moment independence** of order $\boldsymbol{c}$ if for any $\boldsymbol{a} = (a_i)_{i\in[n]} <_{lex} \boldsymbol{c}$, $\boldsymbol{a} \in \mathbb{Z}^n_{\geq 0}$,

$$\mathbb{E}\left(\boldsymbol{\theta}^{\boldsymbol{a}}\right) = \prod_{i\in[n]} \mathbb{E}\left(\theta_i^{a_i}\right),$$

where $\boldsymbol{\theta}^{\boldsymbol{a}} = \theta_1^{a_1} \cdots \theta_n^{a_n}$.

It is generally well known that standard probabilistic independence only guarantees that the first moment of a product can be written as the product of the moments. Separations for higher orders are implied by standard independence only through a cumulant parametrization, where the cumulant generating function

Paper No. 15-08, www.warwick.ac.uk/go/crism

for a product of independent random variables (defined as a random sum of independent realizations) is the composition of the respective cumulant generating functions.

For the purpose of decision support in partial belief systems it is helpful to study moments, since expected utilities often formally depend on these. Now consider for instance two parameters θ_1 and θ_2. Assume a CEU is equal to $\theta_1^2\theta_2^2$ and that a moment independence of order $(2,2)$ holds. Then

$$\mathbb{E}\left(\theta_1^2\theta_2^2\right) = \mathbb{E}\left(\theta_1^2\right)\mathbb{E}\left(\theta_2^2\right) = \mathbb{E}(\theta_1)^2\mathbb{E}(\theta_2)^2 + \mathbb{E}(\theta_1)^2\mathbb{V}(\theta_2) + \mathbb{E}(\theta_2)^2\mathbb{V}(\theta_1) + \mathbb{V}(\theta_1)\mathbb{V}(\theta_2). \tag{5}$$

The same expression is obtained when using sequentially the tower rule of expectations and the law of total variance under the assumption of independence of the two parameters above [2]. Therefore, the expression obtained under moment independence is reasonable and coincides with the one implied by the independence of θ_1 and θ_2. However the condition we need for equation (5) to hold *does not require* θ_1 and θ_2 to be independent.

4. Adequacy in partially defined systems

Given the definitions in Section 3 of new independence concepts tailored for IDSSs, we can now study when adequacy holds.

Theorem 1. *Let $q_d(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d))$ be an algebraic CEU of a quasi independent IDSS. The IDSS is adequate if panel G_i delivers the vectors of expectations $\boldsymbol{\mu}_i(d)$, for all $i \in [m]$ and all $d \in \mathcal{D}$.*

Proof. This result follows by noting that quasi independence implies score separability since

$$\bar{u}(d) = q_d(\mathbb{E}(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d)), \ldots, \mathbb{E}(\boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d))) = \sum_{\boldsymbol{b}\in B} k_{\boldsymbol{b},d} \prod_{i\in[m]} \prod_{j\in[s_i]^0} \mu_{ji}(d).$$

Assuming the CEU is a polynomial in the panels' parameters, under a specific moment independence assumption we have a more operative result.

Corollary 1. *Let $q_d(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d))$ be an algebraic CEU of an IDSS, $\boldsymbol{\theta}_i = (\theta_{ji})_{j\in[s_i]}$ and $\lambda_{ji}(\boldsymbol{\theta}_i, d) = \boldsymbol{\theta}_i^{\boldsymbol{a}_{ji}}$, with $\boldsymbol{a}_{ji} \in \mathbb{Z}_{\geq 0}^{s_i}$, $i \in [m]$, $j \in [s_i]$. Let $\boldsymbol{a}_i^* = (a_{ji}^*)_{j\in[s_i]}$, where a_{ji}^* is the greatest element in $\{a_{ji} : j \in [s_i]\}$, $i \in [m]$, and let $\boldsymbol{a}^* = (\boldsymbol{a}_i^{*\mathrm{T}})_{i\in[m]}$. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_i^{\mathrm{T}})_{i\in[m]}$ and assume the CK-class includes a moment independence assumption of order $\boldsymbol{a}^*$. The IDSS is adequate if panel G_i delivers the vectors of expectations $\boldsymbol{\mu}_i(d)$, for all $i \in [m]$ and all $d \in \mathcal{D}$.*

Proof. Adequacy is guaranteed if the EU function can be written in terms of $\mu_{ji}(d)$ and $k_{\boldsymbol{b},d}$, $i \in [m]$, $j \in [s_i]$ and $d \in \mathcal{D}$. Note that

$$\begin{aligned} \bar{u}(d) &= \mathbb{E}\left(q_d(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, d), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, d))\right) \\ &= \sum_{\boldsymbol{b}\in B} k_{\boldsymbol{b},d}\mathbb{E}\Big(\prod_{i\in[m]} \prod_{j\in[s_i]^0} \lambda_{ji}(\boldsymbol{\theta}_i, d)^{b_{j,i}}\Big) = \sum_{\boldsymbol{b}\in B} k_{\boldsymbol{b},d}\mathbb{E}\Big(\prod_{i\in[m]} \prod_{j\in[s_i]^0} \boldsymbol{\theta}_i^{\boldsymbol{a}_{ji}}\Big). \end{aligned}$$

The argument of this expectation is a monomial of multi-degree lower or equal to $\boldsymbol{a}^*$. Moment independence then implies that

$$\bar{u}(d) = \sum_{\boldsymbol{b}\in B} k_{\boldsymbol{b},d} \prod_{i\in[m]} \prod_{j\in[s_i]^0} \mu_{ji}(d),$$

and the result follows.

Both results start with the assumption of an algebraic CEU. This is often the case in practice [13] and in all the examples below. However, there are families of utility factorizations and statistical models that ensure the associated CEU is algebraic.

Definition 7. Let $\boldsymbol{Y}_i$ be the vector overseen by panel G_i, $i \in [m]$. A utility over $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ is called **panel separable** if it factorizes as

$$u(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m) = \sum_{I \in \mathcal{P}_0([m])} k_I \prod_{i \in I} u_i(\boldsymbol{y}_i),$$

where $\mathcal{P}_0$ denotes the power set without the empty set and k_I is a criterion weight [7].

Definition 8. Under the conditions of Definition 7, a utility over $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ is called **additive panel separable** if it factorizes as

$$u(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m) = \sum_{i \in [m]} k_i u_i(\boldsymbol{y}_i).$$

Under the assumption of an (additive) panel separable utility, each panel can model its preferences over the variables under its jurisdiction using a marginal utility function of its choice. A large class of utilities, often used in practice, are *polynomial* [16]. For simplicity, we assume marginal utility functions to have univariate arguments.

Definition 9. A **polynomial** utility function over y_i of degree $n_i \in \mathbb{Z}_{\geq 1}$ is defined as

$$u(y_i) = \sum_{j \in [n_i]} \rho_{ij} y_i^j,$$

where both the coefficients $\rho_{ij} \in \mathbb{R}$ and the domain of the rewards need to entertain some constraints [7, 16].[1]

The probabilistic model class we consider here is a specific structural equation model (SEM) [26, 27], where each variable is defined through a polynomial function. Henceforth we call these a *polynomial SEM*. SEMs are widely used, especially recently, in the causal literature and a cornerstone reference in the literature is [19].

Definition 10. Let $\boldsymbol{Y} = (Y_i)_{i \in [m]}$ be a random vector. A polynomial structural equation model is defined by

$$Y_i = \sum_{\boldsymbol{a}_i \in A_i} \theta_{i\boldsymbol{a}_i} \boldsymbol{Y}_{[i-1]}^{\boldsymbol{a}_i} + \varepsilon_i, \qquad i \in [m],$$

where $A_i \subset \mathbb{Z}_{\geq 0}^{i-1}$, ε_i is a random error with mean zero and variance ψ_i, $\theta_{i\boldsymbol{a}_i}$ is a parameter, $i \in [m]$, $\boldsymbol{a}_i \in A_i$, and $\boldsymbol{Y}_{[i-1]} = (Y_j)_{j \in [i-1]}$, with $[0] = \emptyset$.

An alternative formulation of the model in Definition 10 in terms of distributions is,

$$Y_i \mid (\boldsymbol{\theta}_i, \boldsymbol{Y}_{[i-1]}) \sim \left(\sum_{\boldsymbol{a}_i \in A_i} \theta_{i\boldsymbol{a}_i} Y_{[i-1]}^{\boldsymbol{a}_i}, \psi_i \right),$$

where $\boldsymbol{\theta} = (\theta_{i\boldsymbol{b}_i})_{\boldsymbol{b}_i \in B_i}$ and $i \in [m]$. These models are suitable candidates for a CK-class since their definition is qualitative in nature and requires only the specification of the relationships between the random variables together with a few selected moments.

For polynomial SEMs and panel separable utilities, the following holds.

Theorem 2. *Assume panel G_i is responsible for Y_i, $i \in [m]$. Assume that the CK-class of an IDSS includes a panel separable utility and a polynomial SEM. Assume each panel agreed to model its marginal utility with a polynomial utility function. Under quasi independence, the IDSS is score separable.*

[1]For simplicity, we assume the intercept to be equal to zero since utilities are unique up to positive affine transformations.

PROOF. Fix a policy $d \in \mathcal{D}$ and suppress this dependence. Under the assumptions of the theorem, the utility function can be written as

$$u(\boldsymbol{y}) = \sum_{I \in \mathcal{P}_0([m])} k_I \sum_{i \in I} \left(\sum_{b_i \in B_i} \rho_{b_i} y_i^{b_i} \right), \tag{6}$$

for $B_i \subset \mathbb{Z}_{>0}$. Note also that we can rewrite (6) as

$$u(\boldsymbol{y}) = \hat{u}(\boldsymbol{y}_{[m-1]}) + \hat{u}(y_m),$$

where

$$\hat{u}(\boldsymbol{y}_{[m-1]}) = \sum_{I \in \mathcal{P}_0([m-1])} k_I \prod_{i \in I} \Big(\sum_{b_i \in B_i} \rho_{b_i} y_i^{b_i} \Big), \quad \hat{u}(y_m) = \sum_{I \in \mathcal{P}_0^m([m])} k_I \prod_{i \in I} \Big(\sum_{b_i \in B_i} \rho_{b_i} y_i^{b_i} \Big), \tag{7}$$

and $\mathcal{P}_0^m([m]) = \mathcal{P}_0([m]) \cap \{m\}$. Calling $\boldsymbol{\theta}$ the overall parameter vector of the IDSS, the CEU function, $\mathbb{E}(u(\boldsymbol{Y}) \mid \boldsymbol{\theta})$, can be written applying sequentially the tower rule of expectation as

$$\mathbb{E}(u(\boldsymbol{Y}) \mid \boldsymbol{\theta}) = \mathbb{E}_{Y_1 \mid \boldsymbol{\theta}} \Big(\cdots \mathbb{E}_{Y_{m-1} \mid \boldsymbol{Y}_{[m-2]}, \boldsymbol{\theta}} \big(\hat{u}(\boldsymbol{y}_{[m-1]}) + \mathbb{E}_{Y_m \mid \boldsymbol{Y}_{[m-1]}, \boldsymbol{\theta}} (\hat{u}(\boldsymbol{y}_m)) \big) \Big). \tag{8}$$

From equation (7), the definition of a polynomial SEM and observing that the power of a polynomial is still a polynomial function of the same arguments, it follows that $E_{Y_m \mid \boldsymbol{Y}_{[m-1]}, \boldsymbol{\theta}} (\hat{u}(\boldsymbol{y}_m)) = p_m(\boldsymbol{Y}_{[m-1]}, \boldsymbol{\theta})$, where p_m is a generic polynomial function. Thus $\hat{u}(\boldsymbol{Y}_{[m-1]}) + \mathbb{E}_{Y_m \mid \boldsymbol{Y}_{[m-1]}, \boldsymbol{\theta}} (\hat{u}(\boldsymbol{y}_m))$ is also a polynomial function of the same arguments. Following the same reasoning, we then have that

$$\mathbb{E}_{Y_{m-1} \mid \boldsymbol{Y}_{[m-2]}, \boldsymbol{\theta}} \Big(\hat{u}(\boldsymbol{y}_{[m-1]}) + \mathbb{E}_{Y_m \mid \boldsymbol{Y}_{[m-1]}, \boldsymbol{\theta}} (\hat{u}(\boldsymbol{y}_m)) \Big) = p_{m-1}(\boldsymbol{Y}_{[m-2]}, \boldsymbol{\theta}),$$

where p_{m-1} is a generic polynomial function. Therefore the same procedure can be applied to all the expectations in (8). So $\mathbb{E}(u(\boldsymbol{Y}) \mid \boldsymbol{\theta}) = p_1(\boldsymbol{\theta})$, where p_1 is a generic polynomial function. This defines by construction an algebraic CEU, where the functions λ_{ij} are monomials. Quasi independence and Lemma 1 then guarantee score separability holds.

Theorem 2 together with Lemma 1 shows that in IDSSs whose CK-class respects the assumptions of the theorem EU scores can be uniquely computed from the individual judgements of the panels. By construction, the quasi independence condition of Theorem 2 actually corresponds to a moment independence. The order of such independence depends on the polynomial form of both the structural equation model and the utility function. In Section 5 we identify the order of the moment independence condition required for adequacy in a subclass of polynomial SEMs.

4.1. Examples

4.1.1. Independence binary models

We begin with a rather simple setting where a small number of summaries are sufficient to determine an EU maximizing decision. Let the CK-class specify that $\boldsymbol{Y} = (Y_i)_{i \in [m]}$, where each variable Y_i is binary and overseen by panel G_i. Assume that for all decisions $d \in \mathcal{D}$, $\theta_i = \mathbb{P}(Y_i = 1 \mid \theta_i, d)$, $\boldsymbol{\theta} = (\theta_i)_{i \in [m]}$, that the CK-class includes the belief that $Y_i \mid (\boldsymbol{\theta}, d)$ are mutually independent. Suppose each panel G_i delivers the set of beta distributions $\mathrm{Be}(p_i, q_i)$ for $\boldsymbol{\theta}_i \mid d$ and that the CK-class includes utility factorizations of the form

$$u(\boldsymbol{y}) = u(y_1, \dots, y_m) = \sum_{i \in [m]} k_i y_i + \sum_{i \in [m]} \sum_{i < j \leq m} k_{ij} y_i y_j.$$

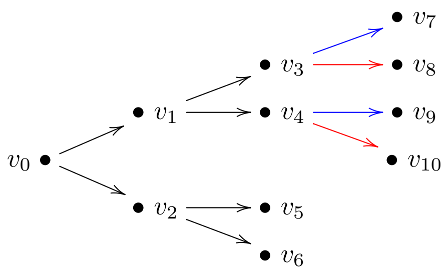

Figure 1: Example of a staged tree model.

With no further assumptions, the CEU can be written as

$$\bar{u}(d \mid \boldsymbol{\theta}) = \sum_{i\in[m]} k_i\lambda_i(\theta_i, d) + \sum_{i\in[m]} \sum_{i<j\leq m} k_{ij}\lambda_i(\theta_i, d)\lambda_j(\theta_j, d), \tag{9}$$

where $\lambda_i(\theta_i, d) = \theta_i$. Thus, equation (9) is an algebraic CEU. In this example quasi independence then corresponds to moment independence of order $\mathbf{1}$, where $\mathbf{1}$ is a vector of dimension m with 1 in all its entries and is implied standard independence. Furthermore, the monomials $\lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, d)$ in equation (3) here become monomials of degree either one or two corresponding respectively to $\lambda_i(\theta_i, d)$ or $\lambda_i(\theta_i, d)\lambda_j(\theta_j, d)$, for $i, j \in [m]$, $j > i$.

Defining $\mu_i = p_i(p_i + q_i)^{-1} = \mathbb{E}(\theta_i \mid d)$ and assuming quasi independence is in the CK-class we obtain that this IDSS is adequate by taking the expectation of equation (9) and

$$\bar{u}(d) = \sum_{i\in[m]} k_i\mu_i + \sum_{i\in[m]} \sum_{i<j\leq m} k_{ij}\mu_i\mu_j.$$

4.1.2. Staged trees

As a second example, we consider now *staged trees* [5, 22], a class of probability trees where certain probabilities are identified. Specifically non leaf vertices of the tree are said to be in the same *stage* if the probabilities associated to their emanating edges are in one-to-one correspondence. For example, v_3 and v_4 of the staged tree in Figure 1 are assumed to lie in the same stage and their identified edges are assigned the same colour. More formally, this tree has four stages $w_0 = \{v_0\}$, $w_1 = \{v_1\}$, $w_2 = \{v_2\}$ and $w_3 = \{v_3, v_4\}$. In [23] we showed that staged trees can be part of a coherent IDSS whenever panels oversee disjoint subsets of its stage set. We suppose that there are three panels G_1, G_2 and G_3 having responsibility over w_0, $\{w_1, w_2\}$ and w_3 respectively. Three binary random variables, Y_l, Y_c and Y_r can be associated to this tree, all with sample space $\{0, 1\}$. We assume the leftmost two edges are the possible outcomes of Y_l, the edges in the center are the outcomes of Y_c given the different levels of Y_l, whilst the rightmost edges coincide with the outcomes of Y_r given $\{Y_c, Y_l\}$.

Assume all the panels have agreed on an additive utility factorization so that

$$u(y_l, y_c, y_r) = k_l u_l(y_l) + k_c u_c(y_c) + k_r u_r(y_r),$$

and that they have been jointly able to further specify the criterion weights. Let $\sigma_{sy} = u_s(y)$, for $s \in \{l, c, r\}$ and $y \in \{0, 1\}$, and θ_{ij} denote the probability of going from v_i to v_j, for every $d \in \mathcal{D}$, $i \in [4]^0$ and $j \in [10]$. Note that the staged tree in Figure 1 introduces the constraints $\theta_{37} = \theta_{49}$ and $\theta_{38} = \theta_{4(10)}$.

Through a sequential application of the tower rule of expectation, it can be easily deduced that the CEU of this problem can be written as

$$\bar{u}(\boldsymbol{\theta} \mid d) = \bar{u}_l + \bar{u}_c + \bar{u}_r,$$

Paper No. 15-08, www.warwick.ac.uk/go/crism

where

$$\bar{u}_l = k_l\sigma_{l1}\theta_{01} + k_l\sigma_{l0}\theta_{02},$$
$$\bar{u}_c = k_c\sigma_{c1}\theta_{13}\theta_{01} + k_c\sigma_{c0}\theta_{14}\theta_{01} + k_c\sigma_{c1}\theta_{25}\theta_{02} + k_c\sigma_{c0}\theta_{26}\theta_{02},$$
$$\bar{u}_3 = k_r\sigma_{r1}\theta_{37}\theta_{13}\theta_{01} + k_r\sigma_{r0}\theta_{38}\theta_{13}\theta_{01} + k_r\sigma_{r1}\theta_{37}\theta_{14}\theta_{01} + k_r\sigma_{r0}\theta_{38}\theta_{14}\theta_{01}.$$

So this CEU is again algebraic. The coefficients $k(\boldsymbol{b}, d)$ of the monomials in equation (3) correspond to the jointly agreed criterion weights, and the unknown functions $\boldsymbol{\lambda}_s(\boldsymbol{\theta}_s, d)$, $s \in \{l, c, r\}$, are

$$\boldsymbol{\lambda}_l(\boldsymbol{\theta}_l, d) = (\theta_{01}, \theta_{02}, \sigma_{l1}\theta_{01}, \sigma_{l0}\theta_{02})^{\mathrm{T}},$$
$$\boldsymbol{\lambda}_c(\boldsymbol{\theta}_c, d) = (\theta_{13}, \theta_{14}, \theta_{25}, \theta_{26}, \sigma_{c1}\theta_{13}, \sigma_{c0}\theta_{14}, \sigma_{c1}\theta_{25}, \sigma_{c0}\theta_{26})^{\mathrm{T}},$$
$$\boldsymbol{\lambda}_r(\boldsymbol{\theta}_r, d) = (\sigma_{r1}\theta_{37}, \sigma_{r0}\theta_{38})^{\mathrm{T}}.$$

Thus, once again these polynomials can be seen to be a simple multilinear function of probabilities delivered by different panels. Under quasi independence, as guaranteed by Lemma 1, an IDSS so defined is adequate.

5. Bayesian networks

Detailed results are valid for the BN model class, where each variable of the network is defined by a specific polynomial SEM introduced in Definition 11.

Definition 11. A BN over a directed acyclic graph (DAG) $\mathcal{G}$ with vertex set $V(\mathcal{G}) = \{i : i \in [m]\}$ and edge set $E(\mathcal{G})$ is a *linear SEM* if each variable Y_i is defined as

$$Y_i = \theta_{0i} + \sum_{j \in \Pi_i} \theta_{ji} Y_j + \varepsilon_i,$$

where Π_i is the parent set of i in $\mathcal{G}$, ε_i is a random error with mean zero and variance ψ_i and $\theta_{0i}, \theta_{ji} \in \mathbb{R}$.

Although such a model is often multivariate Gaussian [25], in general this does not need to be the case.

Just as [25], we consider regression parameters as indeterminates in a polynomial function. We associate these to edges and vertices of the underlying DAG. For $i \in [m]$, let $\theta'_{0i} = \theta_{0i} + \varepsilon_i$ be the indeterminate associated to the vertex i, whilst θ_{ij}, for $(i,j) \in E(\mathcal{G})$.[2] Define $\vec{P}_i$ as the set of rooted directed paths in $\mathcal{G}$ ending in Y_i. A *rooted path* of length $n+1$ from i_1 to j_n is a sequence comprising of a vertex in $V(\mathcal{G})$ and n distinct edges in $E(\mathcal{G})$ is such that $(i_1, (i_1, j_1), \dots, (i_k, j_k), (i_{k+1}, j_{k+1}), \dots, (i_n, j_n))$, where $j_k = i_{k+1}$, $k \in [n-1]$, $i_k, j_k \in [m]$. For every element $P \in \vec{P}_i$ we define $\boldsymbol{\theta}_P$ as

$$\boldsymbol{\theta}_P = \prod_{i \in P} \theta'_{0i} \prod_{(i,j) \in P} \theta_{ij},$$

and, just as [25], we call $\boldsymbol{\theta}_P$ the *path monomial*.

Example 2. Consider the DAG in Figure 2. The set $\vec{P}_3$ is equal to

$$\{(3), (2, (2,3)), (1, (1,3)), (1, (1,2), (2,3))\}, \tag{10}$$

and θ'_{03}, $\theta'_{02}\theta_{23}$, $\theta'_{01}\theta_{13}$ and $\theta'_{01}\theta_{12}\theta_{23}$ are the corresponding path monomials.

We call *algebraic substitution* the process of plugging-in the linear regression definition of a random variable of the DAG into the structural equation definition of the child variable. An example illustrates this process.

[2] We think of θ'_{0i} as a parameter although this consists of the sum of a parameter θ_{01} and an error ε_i. Note however that from a Bayesian viewpoint these are both random variables.

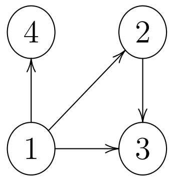

Figure 2: Example of a DAG depicting the relationships between four random variables.

Example 3. For the DAG in Figure 2, the variables of a linear SEM are defined as

$$Y_4 = \theta_{04} + \theta_{14} Y_1 + \varepsilon_1, \qquad Y_3 = \theta_{03} + \theta_{13} Y_1 + \theta_{23} Y_2 + \varepsilon_3,$$
$$Y_2 = \theta_{02} + \theta_{12} Y_1 + \varepsilon_2, \qquad Y_1 = \theta_{01} + \varepsilon_1.$$

An algebraic substitution of the variables in the definition of Y_3 entails

$$\begin{aligned} Y_3 &= \theta_{03} + \theta_{13}(\theta_{01} + \varepsilon_1) + \theta_{23}(\theta_{02} + \theta_{12} Y_1 + \varepsilon_2) + \varepsilon_3 \\ &= \theta_{03'} + \theta_{13}\theta'_{01} + \theta_{23}\theta'_{02} + \theta_{23}\theta_{12} Y_1. \end{aligned}$$

The additional algebraic substitution of Y_1 gives

$$Y_3 = \theta'_{03} + \theta_{13}\theta'_{01} + \theta_{23}\theta'_{02} + \theta_{23}\theta_{12}\theta'_{01}. \tag{11}$$

It is of special interest that after this substitution Y_3 is now uniquely defined in equation (11) in terms of path monomials. Proposition 1 formalizes that this occurs for any variable of a DAG defined as a linear SEM.

Proposition 1. *For a linear SEM over a DAG $\mathcal{G}$, through algebraic substitutions each variable Y_i, $i \in [m]$, can be written as*

$$Y_i = \sum_{P \in \vec{P}_i} \boldsymbol{\theta}_P.$$

Proof. We prove this result via induction over the indices of the variables. Let Y_1 be a root of $\mathcal{G}$. Thus $Y_1 = \theta'_{01}$, where θ'_{01} is the monomial associated to the only rooted path ending in Y_1, namely (Y_1). Assume the result is true for Y_{n-1} and consider Y_n. By the inductive hypothesis we have that, if $i < j$ whenever $i \in \Pi_j$,

$$Y_n = \theta'_{0n} + \sum_{i \in \Pi_n} \theta_{in} Y_i = \theta'_{0n} + \sum_{i \in \Pi_n} \theta_{in} \sum_{P \in \vec{P}_i} \boldsymbol{\theta}_P. \tag{12}$$

Note that every rooted path ending in Y_n is either (Y_n) or consists of a rooted path ending in Y_i, $i \in \Pi_n$, together with the edge (Y_i, Y_n). From this observation the result then follows by rearranging the terms in equation (12).

An algebraic substitution corresponds to computing the conditional expectation of a random variable as formalized by Proposition 2.

Proposition 2. *For a linear SEM over a DAG $\mathcal{G}$, taking $\boldsymbol{\theta}_i = (\theta'_{0i}, \theta_{ji})^{\mathrm{T}}_{j \in \Pi_i}$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_i^{\mathrm{T}})_{i \in [m]}$, $i \in [m]$, we have that*

$$\mathbb{E}(Y_i \mid \boldsymbol{\theta}, d) = \sum_{P \in \vec{P}_i} \boldsymbol{\theta}_P.$$

Proof. This result can be proven via the same inductive process as in the proof of Proposition 1, noting that $\mathbb{E}(Y_1 \mid \boldsymbol{\theta}, d) = \theta'_{01}$ and $\mathbb{E}(Y_n \mid \boldsymbol{\theta}, d) = \theta'_{0n} + \sum_{i \in \Pi_n} \theta_{in} \mathbb{E}(Y_i \mid \boldsymbol{\theta}, d)$.

5.1. Additive factorizations.

Given Propositions 1 and 2, we are now able to write the CEU of polynomial additive panel separable utilities as a polynomial function of a set of monomials readable into the structure of the DAG.

Lemma 2. *Consider a linear SEM over a DAG $\mathcal{G}$. Assume that $u(\boldsymbol{y})$ can be written as*

$$u_i(\boldsymbol{y}) = \sum_{i\in[m]} k_i u_i(y_i).$$

and that u_i is a polynomial utility function of degree n_i. Then the CEU is algebraic and can be written as

$$\bar{u}(d \mid \boldsymbol{\theta}) = \sum_{i\in[m]} k_i \sum_{j\in[n_i]} \rho_{ij} \sum_{|\boldsymbol{\alpha}_i|=j} \binom{j}{\boldsymbol{a}_i} \boldsymbol{\theta}_{\vec{P}_i}^{\boldsymbol{a}_i}, \tag{13}$$

where $\boldsymbol{a}_i = (a_{ij})_{j\in[\#\vec{P}_i]} \in \mathbb{Z}_{\geq 0}^{\#\vec{P}_i}$, $\boldsymbol{\theta}_{\vec{P}_i} = \prod_{P\in\vec{P}_i} \boldsymbol{\theta}_P$, $\binom{j}{\boldsymbol{a}_i}$ is a multinomial coefficient, $\#\vec{P}_i$ is the number of elements in $\vec{P}_i$ and $|\boldsymbol{a}_i| = \sum_{j\in\#\vec{P}_i} a_{ij}$.

Proof. From Proposition 2, it follows that

$$\mathbb{E}(\bar{u}(d \mid \boldsymbol{\theta})) = \sum_{i\in[m]} k_i \sum_{j\in[n_i]} \rho_{ij} \Big(\sum_{P\in\vec{P}_i} \boldsymbol{\theta}_P \Big)^j .$$

The result follows applying the Multinomial Theorem [4].

Equation (13) is an instance of the computation of the moments of a decomposable function as studied in [3] and [17]. In Lemma 2 we explicitly deduce the required monomials and their degree. In the following section we generalise the results in [3] and [17] to generic multilinear functions.

Lemma 2 has an appealing intuitive graphical interpretation which is particularly useful for the computation of the monomials in both the EU in equation (13) and any marginal moment of a linear SEM. The j-th non central moment of any Y_i can be written as the sum of the monomials $\boldsymbol{\theta}_{\vec{P}_i}$ with degree j. By the properties of multinomial coefficients, this sum can be thought of as the sum over the set of unordered j-tuples of rooted paths ending in Y_i. Let $\vec{P}_i^j$ be the set of unordered j-tuples from $\vec{P}_i$. For a $P \in \vec{P}_i^j$, the multinomial coefficient in equation (13) counts the distinct permutations of the elements of P, denoted as n_{P_i}. We then have that

$$\sum_{|\boldsymbol{a}_i|=j} \binom{j}{\boldsymbol{a}_i} \boldsymbol{\theta}_{\vec{P}_i}^{\boldsymbol{a}_i} = \sum_{P\in\vec{P}_i^j} n_{P_i} \prod_{p\in P} \boldsymbol{\theta}_p. \tag{14}$$

Equation (14) becomes thus an intuitive graphical interpretation of equation (13).

Example 4. For the vertex 4 in the DAG of Figure 2 the set $\vec{P}_4$ is equal to $\{(4),(1,(1,4))\}$. From the left hand side of equation (14), Y_4^2 can be written as

$$\theta_{04}'^2 + \theta_{01}'^2\theta_{14}^2 + 2\theta_{01}'\theta_{14}\theta_{04}'. \tag{15}$$

The polynomial above can be equally deduced by simply looking at the DAG. Note that

$$\vec{P}_4^2 = \Big\{ \big((4),(4)\big), \big((1,(1,4)),(1,(1,4))\big), \big((4),(1,(1,4))\big) \Big\}.$$

The first and second monomial in equation (15) correspond to the first and second element of $\vec{P}_4^2$ respectively, whilst the last elements of this set, having two distinct permutation of its elements, is associated to the third monomial in equation (15).

From Lemma 2 we can deduce the independences needed for adequacy in BNs. Note that potentially $\boldsymbol{\theta}_{\vec{P}_i}$ multiplies the same parameter a number of times dependent on the topology of the DAG. We let $\boldsymbol{\theta}_{\mathcal{G}_i}$ be the simplified version of $\boldsymbol{\theta}_{\vec{P}_i}$ where each parameter appears only once and $\boldsymbol{\theta}_{\mathcal{G}_i}^{\boldsymbol{c}_i}$ is the simplified version of $\boldsymbol{\theta}_{\vec{P}_i}^{\boldsymbol{a}_i}$ where each element of $\boldsymbol{c}_i$ equals the sum of the a_{ij} associated to the same parameter. Let l_P be the length of a rooted path P and $l_i = \sum_{P\in\vec{P}_i} l_P$.

Theorem 3. *Suppose the CK-class of an IDSS includes a linear SEM over a DAG $\mathcal{G}$, where panel G_i oversees Y_i, $i \in [m]$ and an additive panel separable utility function. Suppose panel G_i agreed to model its marginal utility with a polynomial utility function of degree $n_i \in \mathbb{Z}_{>0}$, $i \in [n]$. For $\boldsymbol{a} \in \mathbb{Z}_{\geq 0}^{l_i}$, if $\boldsymbol{\theta}_{\mathcal{G}_i}$ entertains moment independence of order $\boldsymbol{c}_i$ for every $|\boldsymbol{c}_i| = n_i$ and $i \in [m]$, then the IDSS is score separable.*

Proof. Under the assumptions of the theorem, the CEU function can be written as in equation (13). From the linearity of the expectation operator we have that

$$\mathbb{E}(\bar{u}(d \mid \boldsymbol{\theta})) = \sum_{i\in[m],j\in[n_i]} \rho_{ij} \sum_{|\boldsymbol{a}_i|=j} \binom{j}{\boldsymbol{a}_i} \mathbb{E}\left(\boldsymbol{\theta}_{\vec{P}_i}^{\boldsymbol{a}_i}\right) = \sum_{i\in[m],j\in[n_i]} \rho_{ij} \sum_{|\boldsymbol{c}_i|=j} \binom{j}{\boldsymbol{c}_i} \mathbb{E}\left(\boldsymbol{\theta}_{\mathcal{G}_i}^{\boldsymbol{c}_i}\right).$$

Applying moment independence and letting V_i and E_i be the sets of distinct vertices and edges, respectively, for all the elements $P \in \vec{P}_i$, we have that

$$\mathbb{E}(\bar{u}(d \mid \boldsymbol{\theta})) = \sum_{\substack{i\in[m],j\in[n_i],\\ |\boldsymbol{c}_i|=j}} k_i\rho_{ij} \binom{j}{\boldsymbol{c}_i} \prod_{l\in V_i} \mathbb{E}\left(\theta_{0l}^{\prime c_{il}} \theta_{lCh_l}^{c_{iCh_l}}\right) \prod_{(j,k)\in E_i\setminus(l,Ch_l)} \mathbb{E}\left(\theta_{jk}^{c_{ik}}\right),$$

where c_{ik} is the element of $\boldsymbol{c}_i$ associated to θ_{jk} and Ch_l is the index of a children of the vertex l. The thesis then follows.

Theorem 3 can be seen as an instance of Theorem 2 where the specific moment independences necessary for the IDSS's adequacy has been made explicit. By requesting the collective to agree on these independences, the IDSS can then quickly produce a unique EU score for each policy. Panels are informed on the summaries that they need to deliver to the IDSS since these are the only quantities of which the EU is a function.

5.2. Multilinear factorizations.

The algebraic approach we have taken in this paper enables us to generalize in a straightforward manner the results in Section 5.1 about additive/decomposable factorizations so they apply to multilinear functions. Let $\#\vec{P}_i = m_i$, $\boldsymbol{l} = (\boldsymbol{l}_i^{\mathrm{T}})_{i\in[m]}$, $\boldsymbol{l}_i = (l_{ij})_{j\in[m_i]} \in \mathbb{Z}_{\geq 0}^{m_i}$ and $\boldsymbol{a} = (a_i)_{i\in[m]} \in \mathbb{Z}^m$. We write $\boldsymbol{l} \simeq \boldsymbol{a}$ if both $|\boldsymbol{a}| = |\boldsymbol{l}|$ and, for all $i \in [m]$, $|\boldsymbol{l}_i| = a_i$.

Lemma 3. *For a linear SEM over a DAG $\mathcal{G}$, suppose the utility function $u(\boldsymbol{y})$ can be written*

$$u(\boldsymbol{y}) = \sum_{I\in\mathcal{P}_0([m])} k_I \prod_{i\in I} u_i(y_i).$$

Suppose u_i is a polynomial utility function of degree $n_i \in \mathbb{Z}_{\geq 1}$, $\boldsymbol{n} = (n_i)_{i\in[m]}$, $i \in [m]$ and $\boldsymbol{0}$ is a vector of dimension m with only zero entries. The CEU is then algebraic and can be written as

$$\bar{u}(d \mid \boldsymbol{\theta}) = \sum_{\boldsymbol{0}<_{lex}\boldsymbol{a}\leq_{lex}\boldsymbol{n}} c_{\boldsymbol{a}} \sum_{\boldsymbol{l}\simeq\boldsymbol{a}} \binom{|\boldsymbol{a}|}{\boldsymbol{l}} \boldsymbol{\theta}_{P_{tot}}^{\boldsymbol{l}}, \tag{16}$$

where $c_{\boldsymbol{a}} = k_J \prod_{j\in J} \rho_{ja_j}$, $J = \{j \in [m] : a_j \neq 0\}$, and $\boldsymbol{\theta}_{P_{tot}} = \prod_{i\in[m]} \boldsymbol{\theta}_{\vec{P}_i}$.

PROOF. To prove this result we first show that under the assumptions of the lemma the utility function can be written as

$$u(\boldsymbol{y}) = \sum_{\mathbf{0} <_{lex} \boldsymbol{a} \leq_{lex} \boldsymbol{n}} c_{\boldsymbol{a}} \boldsymbol{y}^{\boldsymbol{a}}, \tag{17}$$

and then prove that

$$\boldsymbol{Y}^{\boldsymbol{a}} = \sum_{\boldsymbol{l} \simeq \boldsymbol{a}} \binom{|\boldsymbol{a}|}{\boldsymbol{l}} \boldsymbol{\theta}^{\boldsymbol{l}}_{P_{tot}}. \tag{18}$$

The lemma will then follow by substituting into equation (17) for $\boldsymbol{y}^{\boldsymbol{a}}$ given in equation (18).

We prove equation (17) via induction over the number of vertices of the DAG. If the DAG has only one vertex then

$$u(\boldsymbol{y}) = k_1 \sum_{i \in n_1} \rho_{1i} y_1^i.$$

This can be seen as an instance of equation (17). Assume the result holds for a network with $n-1$ vertices. A multilinear utility factorisation can be rewritten as

$$u(\boldsymbol{y}) = \sum_{I \in \mathcal{P}_0([n-1])} k_I \prod_{i \in I} u_i(y_i) + \sum_{I \in \mathcal{P}_0^n([n])} k_I \prod_{i \in I \setminus \{n\}} u_i(y_i) u_n(y_n) + k_n u_n(y_n). \tag{19}$$

The first term on the rhs of (19) is by inductive hypothesis equal to the sum of all the possible monomial of degree $\boldsymbol{a}^{\mathrm{T}} = (a_1, \dots, a_{n-1}, 0)$ where $0 < a_i < n_i$, $i \in [n]$. The other terms only include monomials such that the exponent of y_n is not zero. Letting $\boldsymbol{n}_{i-1} = (n_i)_{i \in [n-1]}$, $\boldsymbol{y}_{[n-1]} = \prod_{i \in [n-1]} y_i$ and $u' = \sum_{I \in \mathcal{P}_0^n([n])} k_I \prod_{i \in I \setminus \{n\}} u_i(y_i) u_n(y_n) + k_n u_n(y_n)$, we now have that

$$\begin{aligned} u' &= \sum_{\mathbf{0} <_{lex} \boldsymbol{a} \leq_{lex} \boldsymbol{n}_{n-1}} c_{\boldsymbol{a}} \boldsymbol{y}^{\boldsymbol{a}}_{[n-1]} \Big(\sum_{i \in [n_n]} \rho_{ni} y_n^i \Big) + k_n u_n(y_n) \\ &= \sum_{\substack{\mathbf{0} <_{lex} \boldsymbol{a} \leq_{lex} \boldsymbol{n}_{n-1} \\ i \in [n_n]}} c_{\boldsymbol{a}} \rho_{ni} \boldsymbol{y}^{\boldsymbol{a}}_{[n-1]} y_n^i + k_n u_n(y_n) = \sum_{\substack{\mathbf{0}' <_{lex} \boldsymbol{a} \leq_{lex} \boldsymbol{n}_n \\ a_n \neq 0}} c_{\boldsymbol{a}} \boldsymbol{y}^{\boldsymbol{a}}_{[n]}. \end{aligned} \tag{20}$$

Therefore, equation (17) follows from equations (19) and (20). To prove equation (18) note that the monomial $\boldsymbol{Y}^{\boldsymbol{a}}$ can be written as

$$\boldsymbol{Y}^{\boldsymbol{\alpha}} = \prod_{i \in [m]} Y_i^{a_i} = \prod_{i \in [m]} \left(\sum_{|\boldsymbol{l}_i| = a_i} \binom{a_i}{\boldsymbol{l}_i} \boldsymbol{\theta}^{\boldsymbol{l}_i}_{\vec{P}_i} \right) = \sum_{\boldsymbol{l} \simeq \boldsymbol{a}} \boldsymbol{\theta}^{\boldsymbol{l}}_{P_{tot}} \prod_{i \in [m]} \binom{a_i}{\boldsymbol{l}_i}.$$

Equation (18) then follows by noting that

$$\prod_{i \in [m]} \binom{a_i}{\boldsymbol{l}_i} = \frac{\prod_{i \in [m]} a_i!}{\prod_{i \in [m]} \prod_{j \in [n_i]} l_{ij}!} = \binom{|\boldsymbol{a}|}{\boldsymbol{l}}.$$

Lemma 3 makes a significant generalization to the theory of the computation of moments in decomposable/additive functions of [3] and [17] so that it applies to multilinear functions of BNs defined as a linear SEM. It is interesting to note that the result we derive abobe is connected to the propagation algorithms first developed in [10] to compute the first two moments of certain chain graphs. Here, focusing only on a certain class of continuous DAG models, we are able to explicitly compute, through algebraic substitution, not only the first two moments, but also any other higher order moment of the distribution associated with the graph.

Using again the properties of multinomial coefficients, we can relate equation (16) to the topology of the graph and its rooted paths. For an $\boldsymbol{a} \in \mathbb{Z}^m_{\geq 0}$, let $\vec{P}_{\boldsymbol{a}} = \times_{a_i \neq 0} \vec{P}_i^{a_i}$, where $\times$ denotes the Cartesian product.

$((2),(2),(4),(4))$
$((1,(1,2)),(2),(4),(4))$
$((1,(1,2)),(1,(1,2)),(4),(4))$
$((2),(2),(1,(1,4)),(4))$
$((1,(1,2)),(2),(1,(1,4)),(4))$
$((1,(1,2)),(1,(1,2)),(1,(1,4)),(4))$
$((2),(2),(1,(1,4)),(1,(1,4)))$
$((1,(1,2)),(2),(1,(1,4)),(1,(1,4)))$
$((1,(1,2)),(1,(1,2)),(1,(1,4)),(1,(1,4)))$

Table 1: Tuples of dimension 4 with two paths ending in Y_2 and two more ending in Y_4 in the graph in Figure 2.

This set consists of the unordered $|\boldsymbol{a}|$-tuples of paths, where in each tuple there are a_i paths ending in Y_i. For each element $P \in \vec{P}_{\boldsymbol{a}}$, let $n_P = \sum_{a_i \neq 0} n_{P_i}$. Then we have that

$$\sum_{\boldsymbol{l} \simeq \boldsymbol{a}} \binom{|\boldsymbol{a}|}{\boldsymbol{l}} \boldsymbol{\theta}^{\boldsymbol{l}}_{P_{tot}} = \sum_{P \in \vec{P}_{\boldsymbol{a}}} n_P \prod_{p \in P} \boldsymbol{\theta}_p.$$

This representation of non-central moments in terms of paths extends the computation of the second central moment of [25] via the trek rule to generic non central moments.

Example 5. Consider $\mathbb{E}(Y_2^2 Y_4^2)$. All distinct tuples of dimension four where two paths end in Y_2 and two in Y_4 are summarized in Table 1. The associated conditional expectation can be written as the following polynomial, where the i-th monomial corresponds to the tuple in the i-th row of Table 1:

$$\bar{u}(d \mid \boldsymbol{\theta}) = \theta'^2_{02}\theta'^2_{04} + 2\theta_{12}\theta'_{02}\theta'^2_{04} + \theta^2_{12}\theta'^2_{04} + 2\theta'^2_{02}\theta_{14}\theta'_{04} + \\ 4\theta_{12}\theta'_{02}\theta_{14}\theta_{04} + 2\theta^2_{12}\theta_{14}\theta'_{04} + \theta'^2_{02}\theta^2_{14} + 2\theta_{12}\theta'_{02}\theta^2_{14} + \theta^2_{12}\theta^2_{14}.$$

Note for example that $\theta_{12}\theta'_{02}\theta'^2_{04}$ has coefficient 2 since the paths (Y_2) and $(Y_1,(Y_1,Y_2))$ can be permuted, whilst $\theta_{12}\theta'_{02}\theta_{14}\theta_{04}$ has coefficient 4 since both pairs of paths (Y_2) and $(Y_1,(Y_1,Y_2))$ and (Y_4) and $(Y_1,(Y_1,Y_4))$ can be permuted.

Just as in the additive case, we are now able to deduce the independences required for score separability of an IDSS defined over a BN. Just as in the additive case, we let $\boldsymbol{\theta}^{\boldsymbol{b}}_{\mathcal{G}}$ be the simplified version of $\boldsymbol{\theta}^{\boldsymbol{a}}_{P_{tot}}$ where parameters only appear once and the exponent are appropriately summed. Let $m_i = \#\vec{P}_i$ and $m_{\mathcal{G}} = \sum_{i \in [m]} m_i$, $\boldsymbol{a}_i = (a_{iP})_{P \in \vec{P}_i}$, $\boldsymbol{a} = (\boldsymbol{a}_i^{\mathrm{T}})_{i \in [m]} \in \mathbb{Z}^{l_{\mathcal{G}}}_{\geq 0}$ and $\boldsymbol{n} = (n_i)_{i \in [m]} \in \mathbb{Z}^m_{\geq 0}$.

Theorem 4. *Suppose that the CK-class of an IDSS includes a linear SEM over a DAG $\mathcal{G}$, where panel G_i oversees Y_i, $i \in [m]$ and a panel separable utility. Suppose panel G_i agreed to model its marginal utility with a polynomial utility function of degree $n_i \in \mathbb{Z}_{>0}$, $i \in [m]$. If, for every $\boldsymbol{b} \simeq \boldsymbol{n}$, $\boldsymbol{\theta}_{\mathcal{G}}$ entertains moment independence of order $\boldsymbol{b}$, then the IDSS is score separable.*

PROOF. Under the conditions of the theorem, the CEU function can be written as in (16). The linearity of the expectation operator than implies that

$$\mathbb{E}(\bar{u}(d \mid \boldsymbol{\theta})) = \sum_{\substack{\mathbf{0} <_{lex} \boldsymbol{a} \leq_{lex} \boldsymbol{n} \\ \boldsymbol{l} \simeq \boldsymbol{a}}} c_{\boldsymbol{a}} \binom{|\boldsymbol{a}|}{\boldsymbol{l}} \mathbb{E}\left(\boldsymbol{\theta}^{\boldsymbol{l}}_{P_{tot}}\right) = \sum_{\substack{\mathbf{0} <_{lex} \boldsymbol{b} \leq_{lex} \boldsymbol{n} \\ \boldsymbol{l} \simeq \boldsymbol{b}}} c_{\boldsymbol{b}} \binom{|\boldsymbol{b}|}{\boldsymbol{l}} \mathbb{E}\left(\boldsymbol{\theta}^{\boldsymbol{l}}_{\mathcal{G}}\right).$$

Applying moment independence and letting V_{tot} and E_{tot} be the sets of distinct vertices and edges, respectively, for all the elements $P \in \vec{P}_{tot} = \cup_{i\in[m]} \vec{P}_i$, we then have that for any $\boldsymbol{l} \simeq \boldsymbol{b}$

$$\mathbb{E}\left(\boldsymbol{\theta}^{\boldsymbol{l}}_{\mathcal{G}}\right) = \prod_{t \in V_{tot}} \mathbb{E}\left(\theta_{0t}^{\prime l_{it}} \theta_{tCh_t}^{l_{iCh_t}}\right) \prod_{(j,k)\in E_{tot}\setminus(t,Ch_t)} \mathbb{E}\left(\theta_{jk}^{l_{ik}}\right).$$

Score separability then follows.

This theorem generalizes Theorem 3 to multilinear utility factorizations and thus to a much larger class of IDSSs. It guarantees adequacy in the case an IDSS embeds complex multilinear utility factorization when the structural consensus includes a linear SEM.

6. Discussion

The framework of IDSSs is capable of supporting decision making in situations where judgements come from different panels of experts having jurisdiction over different aspects of the system. In this paper we have relaxed many of the assumptions guaranteeing coherence in this type of systems [5, 23, 24] by exploiting the polynomial structure of certain statistical models and utility functions.

In particular when the structural consensus includes a BN model, the process of algebraic substitution has proven fundamental in identifying the required summaries and independence relations. We have encouraging results towards a generalization of such recursions in dynamic models, as the *multiregression dynamic model* [20], where expressions for the moments can be deduced in closed form. Furthermore, when each vertex of the BN is no longer a random variable but a random vector (for example when a variable is measured at different geographic location), the theory of *tensors* [15] can be employed to concisely report the associated EU expressions. We plan to develop such a methodology in future work.

Acknowledgements

We acknowledge that Manuele Leonelli has been funded by a scholarship of the Department of Statistics of The University of Warwick, whilst James Q. Smith is partly funded by the EPRSC grant EP/K039628/1.

[1] Blangiardo M, Cameletti M (2015) Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley & Sons, Chichester
[2] Brillinger DR (1969) The calculation of cumulants via conditioning. Ann I Stat Math 21(1):215-218
[3] Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ (1999) Probabilistic networks and expert systems. Springer-Verlag, New York
[4] Cox DA, Little J, O'Shea D (2007) Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra. Springer, New York
[5] Freeman G, Smith JQ (2011) Bayesian MAP model selection of chain event graphs. J Multivariate Anal 102(2):1152-1165
[6] Goldstein M, Rougier, J (2006) Bayes linear calibrated prediction for complex systems. J Am Stat Assoc 101(475):1132-1143
[7] Keeney RL, Raiffa H (1976) Decisions with multiple objectives: preferences and value trade-offs. Cambridge University Press, Cambridge
[8] Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. J Roy Stat Soc B 63(3):425-464
[9] Korb KB, Nicholson AE (2010) Bayesian artificial intelligence. CRC press, Boca Raton
[10] Lauritzen SL (1992) Propagation of probabilities, means and variances in mixed graphical association models. J Am Stat Assoc 87(420):1098-1108
[11] Leonelli M, Smith JQ (2013) Using graphical models and multi-attribute utility theory for probabilistic uncertainty handling in large systems, with application to the nuclear emergency management. In ICDEW pp. 181-192
[12] Leonelli M, Smith JQ (2014) Bayesian decision support for complex systems with many distributed experts. CRISM Res Report 14-15. Invited revision to Ann Oper Res
[13] Madsen AL, Jensen F (2005) Solving linear-quadratic conditional Gaussian influence diagrams. IJAR 38(3):263-282

[14] Madsen AL, Jensen F, Kjaerulff UB, Uffe B, Lang M (2005) The Hugin tool for probabilistic graphical models. IJAIT 14(03): 507-543
[15] McCullagh P (1987) Tensor methods in statistics. Chapman and Hall, London
[16] Müller SM, Machina MJ (1987) Moment preferences and polynomial utility. Econ Lett 23(4):349-353
[17] Nilsson D (2001) The computation of moments of decomposable functions in probabilistic expert systems. In Proceedings of the Third International Symposium on Adaptive Systems pp. 116-121
[18] Papamichail KN, French S (2013) 25 Years of MCDA in nuclear emergency management. IMA J Manag Math 24(4):481-503
[19] Pearl J (2000) Causality: models, reasoning and inference. Cambridge University Press, Cambridge
[20] Queen CM, Smith JQ (1993) Multiregression dynamic models. J Roy Stat Soc B 55(4):849-870
[21] Smith JQ (1996) Plausible Bayesian games. In Bayesian Statistics 5 pp. 387-406
[22] Smith JQ, Anderson PE (2008) Conditional independence and chain event graphs. Artif Intell 172(1):42-68
[23] Smith JQ, Barons MJ, Leonelli M (2015) Coherent frameworks for statistical inference serving integrated decision support systems. CRISM Res Report 15-09.
[24] Spiegelhalter DJ, Lauritzen SL (1990) Sequential updating of conditional probabilities on directed graphical structures. Networks 20(5):579-605
[25] Sullivant S, Talaska K, Draisma J (2010) Trek separation for Gaussian graphical models. Ann Stat 38(3):1665-1685
[26] Ullman JB, Bentler PM (2003) Structural equation modeling. In Handbook of Psychology, Wiley Online Library
[27] Wall MM, Amemiya Y (2000) Estimation for polynomial structural equation models. J Am Stat Assoc 95(451):929-940