Bayesian Decision Support in Complex Modular Systems: an Algebraic and Graphical Approach

by

Manuele Leonelli

Thesis

Submitted to The University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

May 2015

THE UNIVERSITY OF
WARWICK

Contents

Acknowledgements

Declaration

I hereby declare that this thesis is the result of my own work and research, except where otherwise indicated. This thesis has not been submitted for examination to any institution other than the University of Warwick.

Some of this work has been published or is currently under the submission process as follows:

- part of the discussion in Chapter 1 has now appeared in the proceedings of the ICDE2013 conference under the title "*Using graphical models and multi-attribute utility theory for probabilistic uncertainty handling in large systems, with application to the nuclear emergency management*", written jointly with my supervisor Jim Smith, and can be found in the bibliography as Leonelli and Smith [2013a];

- the material in Section 4.4.2 has been presented at the ANS Winter Meeting 2013 and is due to appear in the conference proceedings under the title "*Dynamic uncertainty handling for coherent decision making in nuclear emergency response*", written jointly with my supervisor Jim Smith, and can be found in the bibliography as Leonelli and Smith [2013b];

- Section 5.6.2 summarises the results of the paper "*A differential approach for staged trees*", written jointly with Christiane Görgen and my supervisor Jim Smith, and has been accepted to appear in the proceedings of the ECSQARU2015 conference. The manuscript can be found in the bibliography as Görgen et al. [2015];

- the majority of the material in Chapter 4 is included in a paper at the invited revision stage for the Annals of Operations Research under the title "*Bayesian decision support for complex systems with many distributed experts*", written jointly with my supervisor Jim Smith, and can be found in the bibliography as Leonelli and Smith [2015];

- the results in Sections 2.5 and 5.6.1 have been submitted to the International Journal of Approximate Reasoning under the title "*Using computer algebra to*

symbolically evaluate discrete influence diagrams", written jointly with my supervisor Jim Smith and Eva Riccomagno, and can be found in the bibliography as Leonelli et al. [2015a];

- the material in Sections 5.2-5.5 is included in a manuscript under the title "*The algebra of integrated partial belief systems*", written jointly with Eva Riccomagno and my supervisor Jim Smith, and is about to be submitted. The manuscript can be found in the bibliography as Leonelli et al. [2015b];

- Chapter 3 is based on the manuscript "*Coherent inference for integrating decision support systems*", written jointly with my supervisor Jim Smith and Martine J. Barons, and is about to be submitted. The manuscript can be found in the bibliography as Smith et al. [2015].

I have been lead author of all the papers listed above, with the exception of Görgen et al. [2015], where myself and Christiane Görgen have joint ownership of the results, and Smith et al. [2015], which reports results jointly achieved by the authors. However, what appears in this thesis has been solely written by me.

Abstract

Nowadays decision centres are required to make choices in complex and evolving environments, described through multiple and interdependent processes with many associated measurements. The objective of a real time decision making centre is to agree to a sequence of efficacious countermeasures. To achieve this it is usually necessary to integrate opinions and information from an often diverse set of stakeholders, articulating several competing objectives and knowledge over different domains of expertise. A collection of decision support systems can enhance such an integration, not only ensuring that all relevant evidence systematically informs policy making, but also encouraging the decision centre to exhibit an underlying consistency across all its components and to address the problem as a whole. In this thesis we develop a formal framework, extending standard Bayesian methodology, enabling the judgements and the models of groups of experts to be coherently aggregated in a unique entity. We discuss when and how it is possible to do so and the conditions the group needs to agree upon. We call this framework integrating decision support system. We then develop a variety of methodologies to enhance such an integration, enabling integrating decision support systems to be feasibly used in practice.

The word cloud in the following page summarises my research.

List of Figures

List of Tables

To Lucilla

Chapter 1

Introduction

The activity of decision making pervades everyday life of human beings. For instance, we choose what type of food to have for dinner or what road to take to go to work. However, not all types of decisions that we have to commit to might be as straightforward as the ones above. For instance, deciding whether or not to shut down a nuclear source term due to a malfunctioning on site or committing to a policy concerning climate change to reduce CO_2 emissions at world level requires an accurate reflection about the consequences of the available courses of action. In these situations the Decision Maker (DM) - the individual who has the responsibility, accountability and authority to determine a decision and possibly implement it - needs to make choices in complex and evolving environments, described through multiple and interdependent processes with many associated measurements [Bennet and Bennet, 2009].

Empirical research has shown that humans exhibit, even in very simple situations, a high number of inconsistencies when making decisions [Tversky and Kahneman, 1974]. Our actions can often be said to be *irrational* - according to some canons of rationality that we discuss in more details in the following. Additional research has further shown however that the number of fallacies DMs may incur in can greatly decrease if they are properly supported during the decision making process [see e.g. Bhandari et al., 2008]. Thinking about our everyday life, when we are faced with a decision problem involving elements we are not familiar with, we often seek for the help of a person with the relative expertise in the appropriate field. We say that an individual has expertise, and therefore is an *expert*, if he/she has a comprehensive and authoritative knowledge of a specific field, not possessed by many people [Caley et al., 2014].

Expert Systems (ESs), usually computer-based, aim at simulating the reasoning of experts and therefore at providing the required support to DMs by codifying expert knowledge. These usually use artificial intelligence techniques to process information, learn,

reason and make predictions, thus enhancing the process of decision making. Within the information systems' literature, ESs are considered as a special class of **Decision Support Systems (DSSs)**, usually called intelligent DSSs [Marakas, 2003]. There is a large number of definitions of what a DSS is. Here we follow French et al. [2009] saying that a "DSS is a computer-based system that supports the decision making process, helping DMs to understand the problem before them and to form and explore the implications of their judgements, and hence to make a decision upon understanding". Both DSSs and ESs can provide great benefits to DMs since these combine the knowledge of several experts into a reusable tool that quickly processes information and makes inferences. For the purpose of this thesis we use the two terms interchangeably, but we note that the term ES is more common within the statistical community, whilst DSS appears more frequently in the information systems' literature.

Although DSSs have now been developed and used in practice in a variety of domains, we argue in this thesis that DMs can only be properly supported in *complex* current situations by an **Integrating Decision Support System (IDSS)** [Leonelli and Smith, 2015, Smith et al., 2015], which networks together different DSSs into a unique coherent supporting tool. Here, briefly, we notice that this is true mainly for two reasons. First, current applications, as for example nuclear emergency management, require knowledge on a variety of domains, which can be difficult to synthesise into a unique system. Second, the network structure connecting different DSSs provides the basis for both distributed computations and reasoning, increasing the comprehensibility of the analyses performed and the speed of computations, so that the IDSS can be used in real-time.

In order to support more strongly our statements, and therefore highlight both the relevance and the contributions of this thesis, we need to provide a short history of the development of ESs within the statistical literature. In Section 1.1 we describe early ESs in use in the last century. We take a little digression in Section 1.2 to illustrate the use of ESs for nuclear emergency management, summarising the discussion in Leonelli and Smith [2013a]. In Section 1.3 we describe modern statistical DSSs and discuss how these address the needs of current applications. Section 1.4 highlights the need for and the main features of IDSSs by noticing the flaws of current DSSs in properly supporting the decision making process in complex domains. We then conclude the chapter by specifying the contributions and the structure of the thesis in Sections 1.5 and 1.6 respectively.

1.1   Early Probabilistic Expert Systems

At their inception, ESs were only able to model deterministic domains. Their knowledge base - comprising facts about the world and an inference engine - consisted of a set of deterministic rules and, for this reason, such deterministic systems are often referred to as rule-based ESs. Castillo [1997], whilst including an overview of rule-based systems and their domains of application, recognised that the deterministic assumption is too restrictive in most, if not all, of the cases. As the complexity of the problems addressed by ESs increased, a slow recognition of the need for introducing measures of uncertainty began to be appreciated. Uncertainty, as firmly stated by French [1995], is part of any decision analysis, arising, among the others, from physical randomness, imprecision of meaning and various estimates, and therefore needs to be properly addressed.

However at first, probabilistic reasoning was not considered as a viable option for uncertainty modelling in ESs for a variety of reason: probability theory was thought as epistemologically inadequate in this context, it required the specification of an infeasible number of probabilities and there was no technology to feasibly perform the appropriate computations. For these and other reasons, different approaches to the representation of uncertainty were followed, as for example non-monotonic and fuzzy logic, certainty factors and Dempster-Shafer belief functions [see e.g. Buchanan and Shortliffe, 1984, Gabbay, 1985, Klir and Yuan, 1995, Shafer, 1976]. However, around the 1980s, solutions to the three main criticism of the use of probability in ESs began to appear, which allowed for a wider applicability of probabilistic ESs.

At that time the frequentist interpretation of probability was the most widespread within the statistical community. This school of thought defines probability as the long-run frequency of occurrence in repeated events: probability is an intrinsic property of the system being observed. Clearly, this definition of probability does not easily fit the process of supporting DMs, since decision situations are unique and human interventions change the observed system. The subjective Bayesian definition of probability, which was starting to be appreciated, defines it as the degree of belief about the state of the world under consideration by the observer. This naturally provides an interpretation of probability that can be used for uncertainty modelling in ESs. This is because within the Bayesian framework probability is not any more an intrinsic property of the system, but belongs to the DM and represents her own beliefs about it.

It was also recognised that graphical representations of the relationships between random variables could be underpinned by conditional independences [Dawid, 1979], enabling large dimensional joint probabilities to be written as products of local distributions of smaller dimension, thus requiring less probability specifications. Formal statistical

graphical models then began to be defined, as for example Markov Networks (MNs) and Bayesian Networks (BNs) [see e.g. Lauritzen, 1996, Pearl, 1988, and Sections 2.3.2.2 and 2.3.2.4], which, exploiting conditional independences, represent the qualitative structure of a multivariate random vector through a graph. Techniques exploiting the graph structure, as for instance triangulation [see e.g. Spiegelhalter et al., 1993] and Bayesian stochastic numerical approximations [see e.g. Gilks et al., 1996], enabled probability calculations over the graph to be quickly and feasibly computed by ESs.

Statistical graphical models have now been applied by both statisticians and scientists in a vast variety of domains and their methodology, although still refined, is well established [Aguilera et al., 2011, Jordan, 2004, Korb and Nicholson, 2003, Oliver and Smith, 1990, Smith, 2010, Uusitalo, 2007]. However, as the nuclear emergency management example in the following section highlights, many areas of science still have to recognise the importance and the need for a complete uncertainty handling. The goal of deducing objective scientific conclusions, which does not allow for uncertainty claims, still permeates scientific reasoning [see for example the recent discussion in Rougier and Crucifix, 2014, concerning the continuous development of higher-resolution simulators for predicting climate change]. Symbolically, in Carter and French [2006] uncertainty is said to be discomforting to the parties involved in decision making, whilst in Ahlbrecht et al. [1997] uncertainty is even referred to as disconcerting for DMs. Carter and French [2006] further highlights the far too limited perspective on uncertainty in the ESs developed. In Chapter 4 we show how current DSSs can lead DMs to irrational behaviour due to the lack of complete probability propagations.

1.2   An Overview of Nuclear Emergency Management

In 1986 an explosion at one of the reactors of the Chernobyl nuclear power plant released a radioactive plume into the environment contaminating large areas of the former Soviet Union. To protect people and food stocks, measures were taken by the governments of the affected countries. Different and often conflicting responses were further taken by many European countries after the accident, confusing the public opinion and leading to an ineffective implementation of countermeasures [Papamichail and French, 2013, Walle and Turoff, 2008]. It was therefore recognised the need of a comprehensive response to nuclear emergencies within the European community. The way to achieve such coherent response was identified to be the development of a common comprehensive DSS for off-site emergency management. Several institutes in Europe then started the development of the Real-time On-line DecisiOn Support system (RODOS) for nuclear emergencies, which would provide consistent predictions unperturbed by national

boundaries [Ehrhardt et al., 1993]. We note here that other DSSs were built by different initiatives during the following years, notably ARGOS [Hoe and Müller, 2003] and MOIRA [Monte et al., 2009]. Here we focus our attention on the RODOS system only, which embodies in its architecture the typical features of a nuclear emergency.

Alongside the development of RODOS, in the early 1990s the International Chernobyl Project aimed at exploring the factors that drove decision making on protective measures after the Chernobyl accident. Since many different parties and institutions were involved in the decision making process at the time, the study was organised in five decision conferences (see Chapter 2.6.1) (one for each of the affected countries - Belarus, Russia and Ukraine - and two at all-Union level) where simple MultiCriteria Decision Analysis (MCDA) models were used to explore the preferences and the beliefs of the different parties. The analyses performed during these meetings clearly showed that not only radiation related health effects and use of resources were relevant factors, as usually conceived by standard cost-benefit analyses for radiation protection decisions, but also the effects on the health of stress and the political acceptability of the countermeasures taken. Note that MCDA methods are based on subjective judgements of the DMs and for this reason had not been used beforehand in emergency management [Papamichail and French, 2013], confirming the discomfort of scientists with non-objective analyses mentioned above. The successful recognition by the International Chernobyl Project of additional factors that drove decision making further highlights the necessity of using subjective techniques in decision making.

Therefore, it was decided that MCDA methods had to be included in any operational DSS for nuclear emergency recover as RODOS. Such a DSS would then combine scientific knowledge about the likelihood of different events with the value judgements about these to rank different agreed available policies and both facilitate the exploration and create a deeper understanding of the problem at hand. The conceptual architecture of the DSS RODOS itself mirrors the above description [Caminada et al., 2000], since the system is composed by the following three subsystems:

- the Analysing SubsYstem (ASY) processes incoming data and forecasts the location and the amount of contamination through time;

- the Countermeasure SubsYstem (CSY) suggests countermeasures, verifies their feasibility and computes their benefit according to several criteria;

- the Evaluating SubsYstem (ESY) ranks policies according to their potential overall benefits and provides explanations for the reasons of such a ranking.
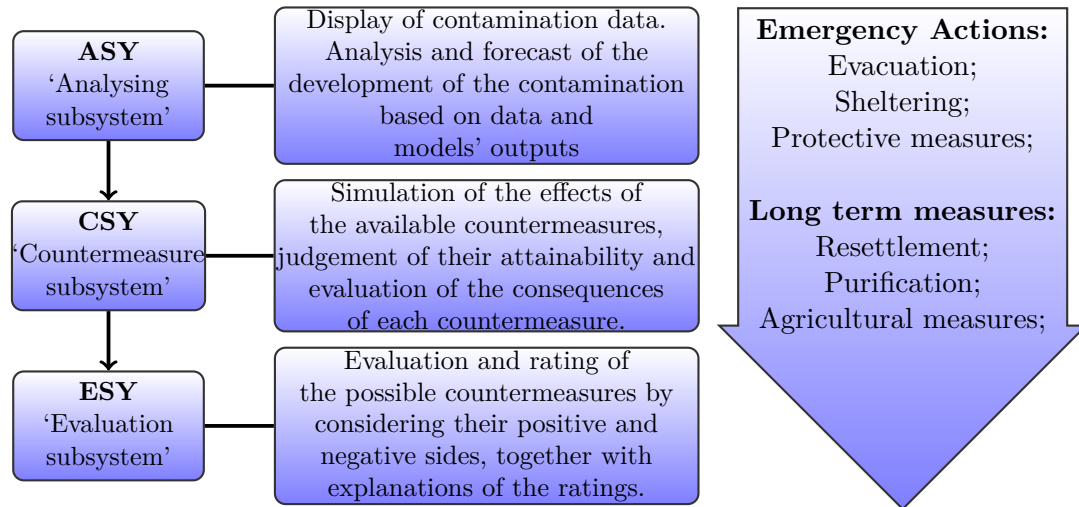
FIGURE 1.1: Conceptual structure of RODOS from Leonelli and Smith [2013a] and after Raskob et al. [2009]. The left column lists the subsystems, the central one describes the functionalities of each of these and the right one summarizes potential countermeasure both in the short and in the long term.

Figure 1.1 summarises how the three subsystems are operationally linked in RODOS. As shown on the left of the diagram, the output of the ASY, consisting of both the likelihood of various events and the forecasts of contamination, are used as input for the CSY. Given these forecasts, the CSY simulates the effects of the available decisions. The results of these simulations are then used by the ESY subsystem to produce a rating of each available countermeasure. We now look into more details into each of these subsystems.

MCDA techniques were included into both the CSY and the ESY components to evaluate and rank potential countermeasures. It is important to note here the relevance of the explanation module within the ESY (as expressed in the bottom-central box of Figure 1.1). Empirical research has shown that DMs do not accept the suggestions of a system which does not provide a rationale for the outputs it produces, even if the system delivers accurate results [Papamichail and French, 2013]. Furthermore, as extensively discussed in French et al. [2009], the ultimate objective of any decision analysis is not in simply choosing a decision which is considered to be optimal according to certain criteria, but more importantly in developing a deeper understanding of the values and the uncertainties of the problem studied. These insights may then lead to an evolution of the DMs' judgements and possibly to a revision of the whole model. This process is often called requisite modelling [Phillips, 1984]. This idea was effectively included in the initial aims of both the development of RODOS and the International Chernobyl Project. Therefore RODOS includes a variety of features to justify its outputs, as for example a natural language report generator to explain the ranking of the policies [Papamichail

and French, 2003], sensitivity analysis graphs and their interpretations [Papamichail and French, 2005].

The ASY subsystem is made of many different *modules* (or component DSSs), each providing estimates and forecasts for a different aspect of the emergency as stated in the top central box of Figure 1.1. For instance, one of these modules concerns the workings of the source term estimating the likelihood of a release of contamination from the plant. Another one includes atmospheric diffusion models describing the spread of the contamination. Additional modules model the effect of this spread might have because of the exposure of humans, animals and plants. Although the ASY modules are capable of working independently, in a comprehensive emergency management these are networked together in the sense that the outputs of some modules are used as inputs for others. Another important element to highlight from the above description is that the domains the ASY modules aim at describing are particularly heterogeneous and therefore require knowledge about a variety of different disciplines. In Figure 1.2, from Leonelli and Smith [2013a], a plausible network of modules for use of a nuclear DSS is presented. Each vertex of the network corresponds to a module and two modules are connected by an arc if the outputs of the parent node are used as inputs for the child node (see Appendix B for an introduction to graph theory). The modules corresponding to the vertices of the network in Figure 1.2 cover the main aspects of a nuclear emergency, as for example the overall workings of a nuclear power plant (Power plant and source term nodes), the spread of contamination (Air and Water dispersal and contamination) and the consequences the accident (Human health, Costs and Political effects). The nodes are grouped in such a way that vertices with the same color represent modules concerning the same domain of expertise. Specifically, the grey vertices are concerned with engineering issues; the green ones with the environment; the blue ones with biological consequences; the brown, red and yellow ones with the political, medical and economical outcomes, respectively. Furthermore, the edge set of the network in Figure 1.2 describes plausible input/output relationships between the modules. So, for instance, the Deposition module uses as input only the outputs of the Water and Air dispersal modules, whilst Human absorption depends on Water dispersal, Animal absorption and Deposition since these are the main factors through which human can get in touch with the contamination.

Each of the modules of the ASY comprises a set of methodologies, either deterministic or stochastic, to perform the task they were designed for. As an example, Turner [1994] lists dozens of different atmospheric dispersion models that have been applied to different domains and systems. From the very early stages of the development of RODOS it was recognised [after considerable debate, as noted in Smith et al., 1997] that stochastic, and more specifically, Bayesian inferential methodologies should be used in such a nuclear DSS to assimilate, combine and represent uncertainties. Smith et al.
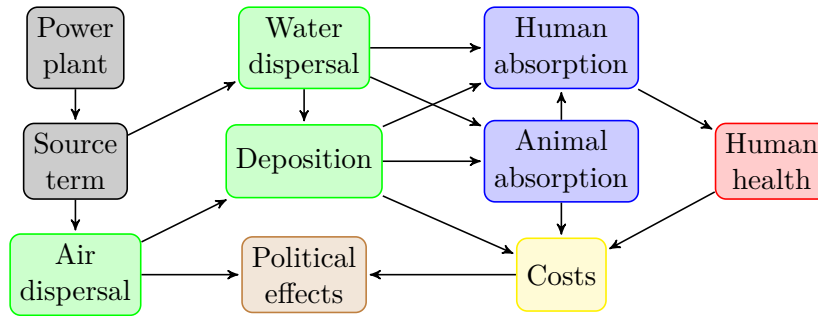
FIGURE 1.2: Plausible network for the modules of a nuclear decision support system.

[1997] and our discussion in Section 1.1 extensively argued why it is reasonable to use this representation of uncertainty. Therefore, different Bayesian models have been developed for the different modules, as for example a BN for the workings of the source term [Caminada et al., 2000], a dynamic random forest for the atmospheric dispersion [Smith et al., 1997] and a dynamic spatial model for the deposition module [De and Faria, 2011].

It is also important to note here that in theory it could be possible for the ASY to consist of a single big module which models the whole domain of nuclear emergency management. However, the provision of the system to be used across all the European countries required the software to support the integration of external programs developed by national research institutes of the many involved countries. Therefore, the ASY had to be structured in this modular form, as a network of sub-models concerning the different elements of the emergency.

The description of the RODOS system for nuclear emergency management has highlighted the complexity of the domains that current DSSs aim at describing. The main points to inherit from this section are the following:

- inference and forecasting need to be distributed among the different modules within the DSS;

- the judgements and beliefs of different individual with different expertise need to be included in the inferential process;

- computations need to be fast to allow for real-time decision making in case of an emergency;

- MCDA techniques need to be included to reveal the true values of the DMs.

It is important to underline the scientific relevance of a project like the development of RODOS. Fortunately, although the consequences of a nuclear accident are and are perceived as severe and far-reaching, the probability of occurrence of these is extremely

low and RODOS is likely not to be used for actual emergency management [Geldermann et al., 2009].[1] At the same time, this implies that data and knowledge about this type of emergencies is scarce. The process of the construction of RODOS has actually deepened the understanding of both the values of potential DMs and the possible evolutions of the emergency. As noted in Papamichail and French [2013], nuclear emergency management has been at the forefront of both stakeholder engagement and the use of decision conferencing, demonstrating the benefits of these techniques. Nowadays RODOS is not only used as an incommensurable educational tool for training personnel through exercises of scenario analysis, but also to highlight the areas in which current emergency management plans may not be adequate. Gering et al. [2013] reported a study where RODOS was used to simulate emergencies with features similar to the recent outworkings of the Fukushima nuclear plant and showed that current methodologies have several major problems.

1.3   Current Probabilistic Expert Systems

The complexity of the models that are currently built to tackle real applications, as the nuclear one, is increasing radically. The decomposition into local distributions that standard graphical models offer is not sufficient any more to both clearly depict the situation under study and allow for fast computations. Within the statistical graphical community, this was first recognised in Mahoney and Laskey [1996] for modelling military problems. The authors highlighted the need of an additional intermediate level of decomposition which breaks down the problem into a set of coupled components, both semantically and formally separable. Semantic separability implies that the component is meaningful to the user of the model, whilst formal separability means that the different components can be re-aggregated into a consistent probability model. This intermediate decomposition exactly corresponds to the division of the ASY component into networked modules within the RODOS system as in Figure 1.2 and is indeed called in Mahoney and Laskey [1996] *modularisation.*

Mahoney and Laskey [1996] referred to two attempts in the literature to create this intermediate level of modelling, namely similarity networks [Heckerman, 1990] and multi-sectioned BNs [Xiang, 2002, Xiang et al., 2011]. However, since both these attempts do not allow for enough modelling flexibility, they pointed towards the use of object oriented techniques in probabilistic modelling, similar to the ones used in programming languages. The authors noticed, as further stressed in Johnson et al. [2012], that the

---

[1]Walle and Turoff [2008] noted that such a system could be actually used for different types of nuclear threats, as for example dirty bombs, although this type of aid is not publicly stated in the mission of RODOS.

object oriented paradigm provides features to simplify the modelling task in complex situations. Specifically:

- **abstraction/encapsulation**: this means that some information is hidden to the user and its access is allowed only via a predefined interface. Modifications within these interfaces do not affect other parts of the system;

- **modularisation/reuse**: the system is composed of a set of loosely connected units, which, if the interfaces are held fixed, can be changed without having to rebuild the whole model.

These features are embedded in the IDSS methodology we introduce in this thesis and, by recalling the description of the nuclear DSS RODOS, are required if we are to develop support systems to be used in practice.

The object oriented route for probabilistic graphical models then started with the development of Object Oriented Bayesian Networks (OOBNs) [Koller and Pfeffer, 1997] and, concerning the modelling of complex situations, then continued in two main streams of research. The first one aimed at integrating first-order logic with Bayesian probability theory and culminated with the definition of multi-entity BNs [Laskey, 2008, Laskey and da Costa, 2009]. These can be represented as a collection MFrags [Laskey and Mahoney, 1997], which can be thought of as OOBNs with root nodes having known values. The other strand of research discusses the use of OOBNs as an integrating tool and introduced the concept of integrating BNs [Johnson and Mengersen, 2012]. An integrating BN can be thought of as a large OOBN where different probabilities are elicited by different groups of experts. These are a special case of IDSSs, but still represent an important step forward to the representation and elicitation of complex probabilistic models. These have been successfully applied to a variety of domains [Johnson et al., 2013, 2014, Mortera et al., 2013].

A completely different route from the object oriented paradigm in ESs has been proposed in Goldstein and O'Hagan [1996], which suggested the use of Bayes linear methodologies not requiring the specification of a full prior distribution [see e.g. Goldstein and Wooff, 2007]. This approach simplifies both the computation and the elicitation burden, but does not provide in itself a viable alternative for the modelling of complex situations as the nuclear emergency management case described above.

Although all the above models can be used to support management decisions as noted in Johnson et al. [2014], these do not include any formal representation of both the available decision space and the preferences of the DM, as required in the nuclear emergency

management example. Influence Diagrams (IDs) are a class of graphical models representing random variables, utilities and decisions [see e.g. Bielza et al., 2011, Howard and Matheson, 2005a, Jensen and Nielsen, 2013, and Section 2.5.1]. Of critical important is to note that the literature about IDs does not provide many insights on the use of these models for the representation and modelling of complex situations. The main focus of ID's research has been on the improvement of the speed computations, either exact or approximated. Thus, although IDs include both decision and utilities, these do not possess the expressive power of, for example, integrating BNs for the representation of current applied problems.

Before concluding this section, we note that the description of the nuclear emergency management domain highlighted the geographic nature of many problems that DMs have to deal with in practice. On one hand, this simply means that the dimensionality of the overall problem grows since each variable of the large system might be recorded and observed at different locations in space. On the other, it also raises issues specific of geographic modelling and concerning the use of geographic information systems in probabilistic models [Stassopoulou et al., 1998]. Although we do not deal specifically with these issues, we recognise their importance in many applications and we note that there are now methods customised for the integration of geographic information systems with graphical models [Johnson et al., 2012, Laskey et al., 2010]. However, in Section 4.4 we present two examples of how a geographic component can be included into an IDSS in the simplest possible case.

1.4 Decision Support Systems for the XXI Century

Although the technology reviewed in the previous section represents a big step forward in the modelling of complex issues, we note here that these methodologies still do not address many of the features of classes of problems as the nuclear emergency management of Section 1.2.

First, because of the underlying OOBN assumption, most of the methods above assume that each submodel is a graphical one. This is a restrictive assumption and, as we saw, many statistical models, not necessarily graphical, have been already developed for some of the ASY components of RODOS. Furthermore some of the modules in use often are not probabilistic: this is the case for example of big simulators that model climate change. The technology of *emulators* [Kennedy and O'Hagan, 2001, O'Hagan, 2006, and Section 2.3.5] can then be used to introduce a probabilistic distribution over the outputs of such simulators.

Second, RODOS implemented MCDA techniques to perform a formal decision analysis. Many current statistical ESs do not allow for such outputs and in general provide only probabilistic beliefs on a set of goal variables. Therefore, there is a need for new methods to be implemented in DSSs that have the power of performing MCDAs. In RODOS this was based on value functions. We argue here though that for current DSSs this is not sufficient and that multiattribute utility theory, enabling a full uncertainty handling, needs to be used as a representation of the DMs' preferences.

Third, in such domains decision making is not usually the responsibility of a single individual but rather of groups. Decision centres will have the accountability and the responsibility of choosing a course of action and their judgements will be supported by best experts' knowledge. Therefore, DSSs need to provide both a theoretical framework allowing for such a collaborative purpose and the technology to support prescriptive team decision making.

Lastly, in order for this to be feasible in practice, both probabilistic inferences and preferential modelling need to be distributed, as in the ASY, ESY, CSY structure of a system like RODOS. Therefore, the distributed nature of, for example, an integrating BN needs to embellished with a formal preferential analysis, which entertains the same distributed nature, both formally and semantically.

It is therefore fundamental to develop a framework with the above features in order to properly support DMs in real current applications. As remarked by Mahoney and Laskey [1996], the distributed nature of such a methodology is vital, since the following benefits would follow:

- computational tractability;

- comprehensibility;

- feasibility of testing.

Computational tractability derives from the local computation structure associated to distributed systems [widely studied in machine learning, see e.g. Peteiro-Barral and Guijarro-Berdiñas, 2013, Rodríguez et al., 2011]. Distributed systems are more comprehensible since the overall outputs can be traced back to each subsytem's outputs, providing much clearer justifications of the delivered estimates. Finally, feasibility of testing follows since each subsystem can be more straightforwardly tested and potentially upgraded than the whole system at once. In the following we highlight how the IDSS methodology we introduce here entertains such features.

As with the motivations that lead to the inception of graphical modelling techniques, we can notice that distributivity in complex systems aims at breaking down a huge problem into ones of smaller dimension. In Spiegelhalter et al. [1993] such a modelling approach was referred to as *divide-and-conquer*. The methodology of IDSSs uses the same idea to simplify decision analyses in highly multivariate and heterogeneous problems. We note here that such an approach is commonly used in a variety of domains and subjects, as for example in composite likelihood methods [Varin et al., 2011], Markov chain Monte Carlo schemes [Lindsten et al., 2014] and parallel computing [Chandy et al., 1998].

1.5 Contributions of the Thesis

In this thesis we develop a methodology that addresses the gaps of current ESs we listed in Section 1.4 to model complex domains. We call our methodology an **IDSS**, which embeds a formal distributed Bayesian decision analysis combining the beliefs of different of groups of experts. Throughout the thesis we specify the conditions enabling such an integration in a variety of frameworks and discuss the advantages of this methodology. We then provide a toolkit of methods for inference and decision making in an IDSS by, for instance, developing propagation algorithms and symbolic methodologies.

From a more specific/domain-based viewpoint, the results in this thesis extend current methods on a variety of aspects:

- in Chapter 3 we extend standard independence notions over the parameter vectors of a variety of graphical models [see e.g. Dawid and Lauritzen, 1993, Freeman and Smith, 2011, Spiegelhalter and Lauritzen, 1990] enabling distributed multi-expert inferences in IDSSs;

- there are now several rules to combine expert judgement in the literature [see e.g. French, 2011]. Our work extends such rules for complex multivariate domains where the overall probability distribution can be described by means of a graphical model (see e.g. Propositions 3.6.5 and 3.6.14);

- a new concept of statistical causality tailored to the needs of an IDSS is defined in Section 3.5. We are able to show that the standard definition of causality of Pearl [2000] represents a special case of our definition;

- we introduce a new class of utility factorisations in Section 4.1.1.3 customised to the needs of an IDSS, allowing for a distributed computation of Bayesian expected utilities in the multi-expert domains we address;

- we introduce in Section 4.3 new utility-based propagation algorithms for the distributed computation of expected utility scores in IDSSs. Some of the current evaluation algorithms of decisional structures can be seen as special cases of the ones we define here;

- we deduce new recursions for the computations of moments of functions over the vertices of an agreed graphical model, generalising the ones of Cowell et al. [1999] and Nilsson [2001];

- by studying decision making in an IDSS from an algebraic perspective, we identify minimal sets of independence statements assuring the system provides rational decision support. This recognition leads us to define in Section 5.3 new types of independence tailored to the needs of an IDSS;

- we develop in Section 5.6 new symbolic approaches to decision making in different frameworks, which extends current symbolic methods for probabilistic reasoning in graphical models [see e.g. Castillo et al., 1997b, Darwiche, 2003].

More detailed explanations of how our results generalise current methodologies can be found in the following chapters.

In this thesis we almost exclusively focus on the technical conditions that guarantee the existence of an IDSS. In order to concisely illustrate this methodology, our examples necessarily need to be small-dimensional. However, as we discuss more extensively in the following chapters, our methods would scale up to larger real-world examples because of the distributed nature of IDSSs.

The material of this thesis has appeared/will appear in seven papers I have co-authored with my supervisor Jim Smith and other collaborators. Additional details about these papers can be found in the Declaration on page v.

1.6 Structure of the Thesis

Having extensively discussed the need for an IDSS and reviewed other statistical approaches to model complex domains in this first chapter, in Chapter 2 we survey classical Bayesian decision analysis methods. We then discuss standard Bayesian reasoning and the Subjective Expected Utility (SEU) methodology for decision making. We introduce a variety of probabilistic, utility and decision models. We further survey issues associated to group decision making and the use of expert judgement.

In Chapter 3 we introduce IDSSs and a set of axioms that can guarantee their existence. We then discuss what makes a 'good' IDSS and generic conditions that can ensure a good IDSS both a priori and after the introduction of both observational and experimental data. We then show that many of the statistical models introduced in Chapter 2 can be used as an underlying probabilistic framework for an IDSS and discuss the conditions that these need to entertain.

We then consider in Chapter 4 a large class of IDSSs based on flexible probability and utility models. For this class we develop distributed utility-based propagation algorithms that allow for the exact computation of expected utility scores from the outputs of the individual modules of the IDSS. We then exemplify the methodology at the end of the chapter in a variety of domains.

The expressions used by the IDSS to rank the available policies are often polynomial functions of the outputs of the component DSSs. This recognition lead us to analyse IDSSs in Chapter 5 from an algebraic viewpoint and to define independence concepts following from the polynomial relationships existing between these outputs. We then introduce symbolic techniques for particular statistical models in IDSSs, which are intimately linked with algebraic analyses.

We conclude the thesis in Chapter 6, where we summarise the main results of the thesis and discuss possible extensions of the methods developed in the thesis.

The thesis also include 5 appendices. In Appendix A we collect longer proofs. Appendix B introduces the required terminology of graph theory. Appendix C reviews basic statistical distributional theory and standard Bayesian models. Appendix D consists of basic definitions of polynomial algebra. In Appendix E we report our computer algebra code to implement the methods we introduce in Section 5.6.1.

Chapter 2

Classical Bayesian Decision Analysis

In the introduction we discussed the use of Bayesian reasoning and modelling techniques at the inception of probabilistic ESs. Although we noted the inappropriateness of standard methodologies for the needs of current applications, several systems are still based on these traditional models. The theory we develop in the following chapters is a generalisation of Bayesian methodologies and models to formally accommodate the beliefs of different group of experts. We need therefore first to provide a broad overview of the currently available Bayesian modelling tools for DMs.

The chapter is structured as follows. In Section 2.1 we outline the Bayesian formalism of inference. Section 2.2 introduces the Subjective Expected Utility (SEU) methodology which embeds canons of rational decision making and defines rules to make decisions that are optimal according to these canons. The SEU model basically consists of two main ingredients: a probability function, representing the beliefs of a DM, and a utility function, representing her values. Sections 2.3 and 2.4 introduce methodologies to represent respectively probability and utility functions in a multivariate setting. In such a case the definition and elicitation of these functions is often prohibitive. Independence concepts have been introduced to factorise these into subfunctions with a smaller number of arguments. Such factorisations can often be depicted by a graph providing a powerful and intuitive representation of the relationships between the elements of the problem. Therefore our discussion is mainly based on graphical models, for both probabilities and utilities. We introduce here a new subclass of certain utility graphical models defined in Abbas [2010] that have interesting properties in the domains we address in this thesis [see also Leonelli and Smith, 2015].

In Section 2.5 we introduce classes of models that explicitly represent both probabilities and utilities, and the decisions available to a single DM. Again our discussion focuses on graphical models and in particular on the Influence Diagram (ID) model. We consider in this thesis the models class of multiplicative IDs defined in Leonelli et al. [2015a] and introduce a new fast evaluation algorithm for these models.

The SEU model is only capable of representing how a single DM should rationally commit to decisions. However, as we have seen in Chapter 1, policy choices are seldom made by just one person. There might be several DMs sharing the authority and responsibility of choosing a certain policy, many stakeholders who are affected by the DM's decision, and experts that are consulted by the DMs. In Section 2.6 we briefly review the literature on both group decision making and the use of expert judgement to support decision making. We then conclude with a more detailed explanation of the contributions of this thesis and a discussion.

2.1 Bayesian Reasoning

2.1.1 The Full Probabilistic Case

A problem domain is denoted by an often large dimensional random vector $\boldsymbol{Y}$ with sample space $\mathcal{Y}$. We denote by $\boldsymbol{y}$ a generic instantiation of $\boldsymbol{Y}$ and we let $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ parametrise the density function f of $\boldsymbol{Y}$. In a Bayesian framework the parameter $\boldsymbol{\theta}$ is itself a random variable. We let π denote its density function. For the purpose of this section, the vectors are assumed to be continuous.[1] Let $\boldsymbol{X}$ be a random vector sampled from the same population as $\boldsymbol{Y}$ providing some data points $\boldsymbol{x}$. Within this framework, **Bayes theorem** can be simply employed to update the beliefs about $\boldsymbol{\theta}$ after evidence $\boldsymbol{x}$ has been observed as

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{f(\boldsymbol{x})} = \frac{f(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} f(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}. \tag{2.1.1}$$

The terms of equation (2.1.1) have all a meaning. The so called *prior distribution*, $\pi(\boldsymbol{\theta})$, represents the beliefs of the DM about the parameter $\boldsymbol{\theta}$ before evidence $\boldsymbol{x}$ is introduced into the system. If for example $\boldsymbol{Y}$ were a univariate binary random variable, then $\pi(\boldsymbol{\theta})$ would simply be a probability distribution over $[0, 1]$ representing the beliefs of the DM about the probability of a success (for simplicity here we have chosen a discrete example). Although we do not further discuss these issues, we note that there is now an extensive literature and several software that allow for a faithful elicitation of such

---

[1]In discrete domains integrals are simply substituted by summations. We denote probability mass functions by p.

prior beliefs [see e.g. O'Hagan et al., 2006]. The other term at the numerator, $f(\boldsymbol{x} \mid \boldsymbol{\theta})$, is usually called a *likelihood function.* In a Bayesian setting this measures the DM's degree of belief in the data taking certain values given that the parameter is fixed to a certain hypothetical value. Continuing the example above, the likelihood would be in this situation the density of a Bernoulli random variable with a hypothetical unknown probability of success. The denominator of equation (2.1.1), $f(\boldsymbol{x})$, simply corresponds to a normalising constant, since for a Bayesian all the relevant information is related to $\boldsymbol{\theta}$. Therefore, equation (2.1.1) is often presented in the following form,

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \propto f(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}), \qquad (2.1.2)$$

where $\propto$ corresponds to proportionality. The corresponding normalising constant is often difficult to compute, since it requires an integration over an arbitrarily large space. Thus, equation (2.1.2) often represents a computationally less intensive version of equation (2.1.1). Finally, the *posterior* distribution, $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$, is simply the revised density function of $\boldsymbol{\theta}$ after evidence $\boldsymbol{x}$ has been observed. Suppose, for instance, that $\boldsymbol{x}$ in the running example is the result of ten coin tosses showing all head (or 1) and assume that the prior distribution was symmetric around $1/2$. The posterior distribution would then be left-skewed and values in the interval $(1/2, 1]$ would have in general higher probability than the ones in $[0, 1/2]$. Henceforth, we assume all these densities to represent the subjective Bayesian beliefs of a DM [for more details on the various interpretation of probabilities and the adequateness of the Bayesian interpretation in decision making, see e.g. French and Rios Insua, 2000, O'Hagan and Forster, 2004].

Depending on whether the variables are continuous or discrete, on their support and on the domain under study, there exist many different possible choices for both the likelihood function and the prior density. We review a variety of standard models in Appendix C, as for example the Bayesian Normal linear model, which assumes a Normal distribution for both the likelihood and the prior density of the mean when the variance is assumed to be known. We note here that a computationally efficient choice of prior and likelihood are of two that are *conjugate* [Bernardo and Smith, 2009, O'Hagan and Forster, 2004]. This is the case whenever the posterior, computed via Bayes theorem, lies in the same family of distributions as the prior. In the example above, if $\pi(\boldsymbol{\theta})$ were chosen to follow a Beta distribution with parameters $(a, b) \in \mathbb{R}^2_{>0}$, then the posterior would also be a Beta with some new parameters $(a', b') \in \mathbb{R}^2_{>0}$ (see Appendix C.1.1).

Implicitly in equation (2.1.1) we assumed that the parameters of the distribution of $\boldsymbol{\theta}$ were known. Within the Bayesian approach this does not necessarily have to be the case and several different levels or layers of uncertainty can be introduced. This modelling strategy is usually referred to as multilevel or *hierarchical modelling* [Gelman and Hill,

2006]. For example, the parameters of the Beta distribution above, a and b, could have been assumed to be random and depending on some other parameters, which could have further been random variables, and so forth. For the purpose of this thesis, we only consider hierarchies consisting of one layer so that the parameters of $\pi(\boldsymbol{\theta})$ are known. These are usually referred to as *hyperparameters*.

2.1.2 Moments

Density functions can be easily summarised using the notion of *moments*, which can describe some of their features. Moments are expected values of power functions of random variables. To introduce these, we first define the concept of expected values. Let $g : \boldsymbol{\mathcal{Y}} \to \boldsymbol{\mathcal{Y}}'$, i.e. g is a function mapping elements of $\boldsymbol{\mathcal{Y}}$ into $\boldsymbol{\mathcal{Y}}'$.

Definition 2.1.1. *The expected value of $g(\boldsymbol{Y})$ is defined as*[2]

$$\mathbb{E}(g(\boldsymbol{Y})) = \int_{\boldsymbol{\mathcal{Y}}'} g(\boldsymbol{y}) f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} = \int_{\boldsymbol{\Theta}} \mathbb{E}(g(\boldsymbol{Y}) \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}, \qquad (2.1.3)$$

where the expected value $\mathbb{E}(g(\boldsymbol{Y}) \mid \boldsymbol{\theta})$ conditional on $\boldsymbol{\theta}$ is defined as

$$\mathbb{E}(g(\boldsymbol{Y}) \mid \boldsymbol{\theta}) = \int_{\boldsymbol{\mathcal{Y}}'} g(\boldsymbol{y}) f(\boldsymbol{y} \mid \boldsymbol{\theta}) \mathrm{d}\boldsymbol{y}. \qquad (2.1.4)$$

For continuity with the previous section, we considered in Definition 2.1.1 quantities conditional on the parameter vector $\boldsymbol{\theta}$. However, in general, the conditioning might operate over any random vector of interest.

We can now define the notion of moment. Let, for $n \in \mathbb{Z}_{\geq 1}$, $[n] = \{1, \ldots, n\}$, $\boldsymbol{s}^{\mathrm{T}} = (s_1, \ldots, s_n) = (s_i)_{i \in [n]} \in \mathbb{Z}_{\geq 0}^n$ and $\boldsymbol{Y^s} = Y_1^{s_1} \cdots Y_n^{s_n}$.

Definition 2.1.2. *We say that the moment of* order $\boldsymbol{s}$ *of $\boldsymbol{Y}$ is $\mathbb{E}(\boldsymbol{Y^s})$. The moment of order $\boldsymbol{s}$ of $\boldsymbol{Y}$ conditional on $\boldsymbol{\theta}$ is $\mathbb{E}(\boldsymbol{Y^s} \mid \boldsymbol{\theta})$.*

We now define some specific moments.

Definition 2.1.3. *Let $\boldsymbol{Y}$, $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$ be random vectors of dimension n. The* expectation *of $\boldsymbol{Y}$ is $\mathbb{E}(\boldsymbol{Y}) = \mathbb{E}(\boldsymbol{Y^1})$, where $\boldsymbol{1}$ is a vector of dimension n with 1 in each entry, whilst the* variance *of $\boldsymbol{Y}$, $\mathbb{V}(\boldsymbol{Y})$, is $\mathbb{E}((\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y}))(\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y}))^{\mathrm{T}})$. The* covariance *between $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$, $\mathbb{C}(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$, is $\mathbb{E}((\boldsymbol{Y}_i - \mathbb{E}(\boldsymbol{Y}_i))(\boldsymbol{Y}_j - \mathbb{E}(\boldsymbol{Y}_j))^{\mathrm{T}})$.*

Definition 2.1.4. *Let $\boldsymbol{Y}$, $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$ be random vectors of dimension n. The* conditional expectation *of $\boldsymbol{Y}$ given $\boldsymbol{\theta}$ is $\mathbb{E}(\boldsymbol{Y} \mid \boldsymbol{\theta}) = \mathbb{E}(\boldsymbol{Y^1} \mid \boldsymbol{\theta})$, whilst the* conditional variance *of $\boldsymbol{Y}$*

---

[2]In the discrete case equations (2.1.3) and (2.1.4) have a slightly different form which is not relevant for this thesis and can be found in Casella and Berger [2002].

given $\boldsymbol{\theta}$, $\mathbb{V}\left(\boldsymbol{Y} \mid \boldsymbol{\theta}\right)$, *is* $\mathbb{E}\left(\left(\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y} \mid \boldsymbol{\theta})\right)\left(\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y} \mid \boldsymbol{\theta})\right)^{\mathrm{T}} \mid \boldsymbol{\theta}\right)$. *The* conditional covariance *between* $\boldsymbol{Y}_i$ *and* $\boldsymbol{Y}_j$ *given* $\boldsymbol{\theta}$, $\mathbb{C}(\boldsymbol{Y}_i, \boldsymbol{Y}_j \mid \boldsymbol{\theta})$, *is* $\mathbb{E}((\boldsymbol{Y}_i - \mathbb{E}(\boldsymbol{Y}_i \mid \boldsymbol{\theta}))(\boldsymbol{Y}_j - \mathbb{E}(\boldsymbol{Y}_j \mid \boldsymbol{\theta}))^{\mathrm{T}} \mid \boldsymbol{\theta})$.

We now list some properties of these operators [see e.g. Casella and Berger, 2002].

Proposition 2.1.5. *Let $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$ be random vectors of dimension n. Further let B_i and B_j be $n \times n$ matrices with real entries, whilst $\boldsymbol{b} \in \mathbb{R}^n$. The following identities hold.*

$$\mathbb{E}(B_i \boldsymbol{Y}_i \pm B_j \boldsymbol{Y}_j \pm \boldsymbol{b} \mid \boldsymbol{\theta}) = B_i \mathbb{E}(\boldsymbol{Y}_i \mid \boldsymbol{\theta}) \pm B_j \mathbb{E}(\boldsymbol{Y}_j \mid \boldsymbol{\theta}) \pm \boldsymbol{b},$$
$$\mathbb{C}(\boldsymbol{Y}_i, \boldsymbol{Y}_j \mid \boldsymbol{\theta}) = \mathbb{E}\left(\boldsymbol{Y}_i \boldsymbol{Y}_j^{\mathrm{T}} \mid \boldsymbol{\theta}\right) - \mathbb{E}(\boldsymbol{Y}_i \mid \boldsymbol{\theta})\mathbb{E}(\boldsymbol{Y}_j \mid \boldsymbol{\theta})^{\mathrm{T}},$$
$$\mathbb{V}(B_i \boldsymbol{Y}_i \pm B_j \boldsymbol{Y}_j \pm \boldsymbol{b} \mid \boldsymbol{\theta}) = B_i \mathbb{V}(\boldsymbol{Y}_i \mid \boldsymbol{\theta})B_i^{\mathrm{T}} + B_j \mathbb{V}(\boldsymbol{Y}_j \mid \boldsymbol{\theta})B_j^{\mathrm{T}} \pm 2B_i \mathbb{C}(\boldsymbol{Y}_i, \boldsymbol{Y}_j \mid \boldsymbol{\theta})B_j.$$

The following proposition then links conditional and unconditional moments of order one and two.

Proposition 2.1.6. *Let $\boldsymbol{Y}_i$ and $\boldsymbol{Y}_j$ be two random vectors. Then*

$$\mathbb{E}(\boldsymbol{Y}_i) = \mathbb{E}(\mathbb{E}(\boldsymbol{Y}_i \mid \boldsymbol{\theta})), \tag{2.1.5}$$
$$\mathbb{V}(\boldsymbol{Y}_i) = \mathbb{E}(\mathbb{V}(\boldsymbol{Y}_i \mid \boldsymbol{\theta})) + \mathbb{V}(\mathbb{E}(\boldsymbol{Y}_i \mid \boldsymbol{\theta})). \tag{2.1.6}$$

Equations (2.1.5) and (2.1.6) are usually called *tower rules* of moments. Specifically, equation (2.1.5) is the tower rule of expectations, whilst equations (2.1.6) is referred to as *law of total variance*. Brillinger [1969] generalised the identities (2.1.5)-(2.1.6) and deduced a recursive formula to compute cumulants of any finite order. Cumulants can be thought of as a function of the moments, and vice versa. Many properties of random vectors can be more easily investigated using cumulants than moments [see e.g. Zwiernik and Smith, 2012].

2.2 The Subjective Expected Utility Model

In the previous section we introduced the Bayesian framework for reasoning about and modelling uncertainty. However, the Bayesian approach further provides an axiomatic basis for decision making, which embodies canons of rationality, usually referred to as the **SEU** model. Broadly speaking this consists of three main components:

- a decision space $\mathcal{D}$ which includes the decisions $\boldsymbol{d} \in \mathcal{D}$ available to the DM;

- a probability density f over the unknown state $\boldsymbol{y} \in \mathcal{Y}$ of a vector of relevant random variables $\boldsymbol{Y}$;

- a utility function $u(\boldsymbol{d}, \boldsymbol{r}(\boldsymbol{y}, \boldsymbol{d}))$ describing the DM's preferences over some random vector $\boldsymbol{R}$, whose instantiation $\boldsymbol{r} = \boldsymbol{r}(\boldsymbol{y}, \boldsymbol{d})$ is a function of both $\boldsymbol{d}$ and $\boldsymbol{y}$, and takes values in $\boldsymbol{\mathcal{R}}$.

Having dealt with probability in Section 2.1, we now focus on utility functions.

Definition 2.2.1. *A utility function with arguments $\boldsymbol{d}$ and $\boldsymbol{r}$ is a real-valued function unique up to increasing affine transformations, $u : \boldsymbol{\mathcal{D}} \times \boldsymbol{\mathcal{R}} \to \mathbb{R}$, such that*

$$\forall\, (\boldsymbol{d}_i, \boldsymbol{r}_i), (\boldsymbol{d}_j, \boldsymbol{r}_j) \in \boldsymbol{\mathcal{D}} \times \boldsymbol{\mathcal{R}}, \;\; u(\boldsymbol{d}_i, \boldsymbol{r}_i) \geq u(\boldsymbol{d}_j, \boldsymbol{r}_j) \Leftrightarrow (\boldsymbol{d}_i, \boldsymbol{r}_i) \succeq (\boldsymbol{d}_j, \boldsymbol{r}_j),$$

where $(\boldsymbol{d}_i, \boldsymbol{r}_i) \succeq (\boldsymbol{d}_j, \boldsymbol{r}_j)$ means that the DM finds that $(\boldsymbol{d}_i, \boldsymbol{r}_i)$ is at least as preferable as $(\boldsymbol{d}_j, \boldsymbol{r}_j)$. The $\times$ denotes the Cartesian product.

Note that in general utility functions only represent preference rankings and do not include any representation of the strength of these preferences. Formally, utilities are only unique up to increasing affine transformations. Therefore, utility functions are usually normalised so that they take values between 0 and 1, i.e. $u : \boldsymbol{\mathcal{D}} \times \boldsymbol{\mathcal{R}} \to [0, 1]$. There have been some theoretical studies considering strength of preferences [see e.g. Argyris et al., 2014], but these are not considered in this thesis.

The study of utility functions dates back to the work of Von Neumann and Morgenstern [1947] which mostly concerned the theory of games. Throughout the following years, several different authors derived different constructions of the SEU model based on different, but related, sets of axioms [e.g. Anscombe and Aumann, 1963, de Groot, 1970, Savage, 1972]. We assume throughout the thesis that a set of axioms underpinning the SEU model holds and that the DM is able to elicit both a utility function and a probability distribution respecting one of these.

Now that utility functions have been introduced, we can show how the SEU model combines probabilities and utilities to guide decision making. Decisions are ranked according to their associated expected utility.

Definition 2.2.2. *Let $\boldsymbol{\mathcal{D}}$ be a decision space, f a density function for a random vector $\boldsymbol{Y}$ parametrised by $\boldsymbol{\theta}$ and u a utility function with arguments $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$ and $\boldsymbol{r} = \boldsymbol{r}(\boldsymbol{y}, \boldsymbol{d})$. The **expected utility** of a decision $\boldsymbol{d}$, $\bar{u}(\boldsymbol{d})$, is defined as*

$$\bar{u}(\boldsymbol{d}) = \int_{\boldsymbol{\Theta}} \bar{u}(\boldsymbol{d} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \boldsymbol{d}) \mathrm{d}\boldsymbol{\theta}, \tag{2.2.1}$$

where

$$\bar{u}(\boldsymbol{d} \mid \boldsymbol{\theta}) = \int_{\boldsymbol{\mathcal{Y}}} u(\boldsymbol{r}, \boldsymbol{d}) f(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{d}) \mathrm{d}\boldsymbol{y}, \tag{2.2.2}$$

*is the **conditional expected utility***.

Definition 2.2.3. *We say that a decision $\boldsymbol{d}^* \in \mathcal{D}$ is **optimal** if $\boldsymbol{d}^*$ maximises the expected utility, i.e.*

$$\boldsymbol{d}^* = \arg\max_{\boldsymbol{d} \in \mathcal{D}} \bar{u}(\boldsymbol{d}). \tag{2.2.3}$$

A few important points need to be made here. First, we underline that the SEU describes rational decision making of a single DM and there is no clear extension of this model to groups of DMs (see Section 2.6). Second, the utility function, in contrast to value functions considered for example within RODOS, are able by construction to model attitudes towards risk [see e.g. Keeney and Raiffa, 1976]. Lastly, we remark that the SEU model falls within the class of normative approaches to decision making. From a set of axioms describing rationality, normative approaches are then able to produce a ranking of the available actions. However, there is a huge body of empirical literature which shows that people do not make decisions according to the SEU model [Tversky and Kahneman, 1974]. This observation led on one hand to the development of extensions of the model [e.g. prospect theory, Kahneman and Tversky, 1979], and on the other to a different use of normative models to support DMs instead of automatically selecting optimal decisions. For example, the SEU model can be implemented into a DSS simply in order to allow the DM to explore the effects of her judgements on the ranking of the actions available to her.

Although equations (2.2.1) and (2.2.2) easily describes how expected utility can be computed, we note that in multivariate and dynamic settings the elicitation of both the utility function and the density function is often hard. Therefore, several concepts of independence and different modelling strategies have been discussed in the literature. We first review the methods developed for the density f and after for the utility function u. Another difficulty arises because of the maximisation in equation (2.2.3) when $\mathcal{D}$ is high-dimensional. We show in Section 2.5 methods that decompose the maximisation process to take place into smaller spaces and therefore be computationally simpler.

2.3 Probability Models

In this section we introduce a variety of models that decompose a probability density f into smaller dimensional (conditional) densities. Our discussion focuses on graphical models since these can more easily describe the qualitative features underlying a domain of interest, a property that turns out to be fundamental for the developments of this thesis. We first introduce in Section 2.3.1 the notion of conditional independence which is at the basis of the decomposition generated by graphical models. We then review

non-dynamic and dynamic models in Sections 2.3.2 and 2.3.3 respectively. In Section 2.3.4 we introduce the notion of statistical causality and explain its importance within Bayesian reasoning. We conclude with a discussion in Section 2.3.5 of the use probabilistic emulators. These describe through a probability density the outputs of a complex deterministic function arising for example from complex computer codes.

For ease of notation we leave the dependence on the decision $\boldsymbol{d}$ implicit in this section. However, although not explicitly labelled, the results below need to hold for every possible decision in $\boldsymbol{\mathcal{D}}$.

2.3.1 Conditional Independence

As discussed in Chapter 1, graphical models can be underpinned by a conditional independence structure [Dawid, 1979] allowing for a formal modularisation of a model. Although conditional independence can be generally thought of as a tertiary operator - · ⫫ · | · - exhibiting a set of properties [see e.g. Dawid and Lauritzen, 1993, Smith, 1989a], we formally define it for random variables and vectors only. It is important to point out that recently Dawid and Constantinou [2014] generalised the notion of conditional independence to include both non-stochastic and stochastic variables and called it *extended conditional independence*, which entertains the same properties of of standard conditional independence.

For the purpose of this section, let A, B C and D be disjoint subsets of $[n]$, $\boldsymbol{Y}_A$, $\boldsymbol{Y}_B$, $\boldsymbol{Y}_C$ and $\boldsymbol{Y}_D$ be random vectors, and g denote a generic function.

Definition 2.3.1. *We say that $\boldsymbol{Y}_A$ is independent $\boldsymbol{Y}_B$ given $\boldsymbol{Y}_C$, $\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C$, if* $f(\boldsymbol{y}_A, \boldsymbol{y}_B, \boldsymbol{y}_C) = f(\boldsymbol{y}_A \mid \boldsymbol{y}_C) f(\boldsymbol{y}_B, \boldsymbol{y}_C)$.

The conditional independence statement $\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C$ asserts that the only information to infer $\boldsymbol{Y}_A$ from $\boldsymbol{Y}_B$ and $\boldsymbol{Y}_C$ is from $\boldsymbol{Y}_C$. Dawid [1979] showed that the conditional independence operator has the following properties.

Proposition 2.3.2. *It holds that:*

1. *$\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C \iff \boldsymbol{Y}_B \perp\!\!\!\perp \boldsymbol{Y}_A \mid \boldsymbol{Y}_C$;*

2. *$\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C \implies g(\boldsymbol{Y}_A) \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C$;*

3. *$\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C \implies \boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C, g(\boldsymbol{Y}_A)$;*

4. *$\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C$ and $\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_D \mid \boldsymbol{Y}_B, \boldsymbol{Y}_C \iff \boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_D, \boldsymbol{Y}_B \mid \boldsymbol{Y}_C$.*

We refer throughout the thesis to the condition in point 1 as the *symmetry property*, whilst point 4 is usually referred to as *strong decomposition*. These properties of conditional independence are usually called *semi-graphoid axioms*.

Note that, although conditional independence is deeply linked to particular factorisations of densities, it is qualitative in nature and therefore easier to identify and justify.

2.3.1.1 Context Specific Independence. The notion of conditional independence is only able to describe statements of the form $\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C$ holding for every instantiation $\boldsymbol{y}_C \in \boldsymbol{\mathcal{Y}}_C$, where $\boldsymbol{\mathcal{Y}}_C$ is the sample space of $\boldsymbol{Y}_C$. However, often this does not have to be the case and a domain can exhibit some independences only for subspaces of the sample space. *Context specific* independences are able to depict this more general scenario [Boutilier et al., 1996].

Definition 2.3.3. *Let $\boldsymbol{\mathcal{Y}}'_D$ be a subspace of $\boldsymbol{\mathcal{Y}}_D$, the sample space of $\boldsymbol{Y}_D$. We say that $\boldsymbol{Y}_A$ and $\boldsymbol{Y}_B$ are contextually independent given $\boldsymbol{Y}_C$ and context $\boldsymbol{y}_D \in \boldsymbol{\mathcal{Y}}'_D$ if*

$$f(\boldsymbol{y}_A, \boldsymbol{y}_B, \boldsymbol{y}_C, \boldsymbol{y}_D) = f(\boldsymbol{y}_A \mid \boldsymbol{y}_C, \boldsymbol{y}_D) f(\boldsymbol{y}_B, \boldsymbol{y}_C, \boldsymbol{y}_D), \qquad \text{for } \boldsymbol{y}_D \in \boldsymbol{\mathcal{Y}}'_D.$$

In Sections 2.3.2.6 and 2.3.2.7 we introduce two classes of models that can graphically model this more general class of conditional independence statements.

2.3.2 Non Dynamic Models

The notions of conditional independence introduced in Section 2.3.1 is now exploited to define various types of models in a non-dynamic framework, i.e. when there is no recursive updating through time of the involved probabilities. Let $\boldsymbol{Y} = (Y_i)_{i \in [n]}^{\mathrm{T}}$ be a vector of random variables with sample space $\boldsymbol{\mathcal{Y}} = \times_{i \in [n]} \mathcal{Y}_i$, where $\mathcal{Y}_i$ is the sample space of Y_i, $\boldsymbol{y} \in \boldsymbol{\mathcal{Y}}$ and $y_i \in \mathcal{Y}_i$, $i \in [n]$. Let $\boldsymbol{\theta}_A$, taking values in $\boldsymbol{\Theta}_A$, parametrise the (conditional) density of $\boldsymbol{Y}_A$ having sample space $\boldsymbol{\mathcal{Y}}_A = \times_{i \in A} \mathcal{Y}_i$, $A \subseteq [n]$. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_i^{\mathrm{T}})_{i \in [n]}^{\mathrm{T}}$ be the parameter associated to the density of $\boldsymbol{Y}$ taking values in $\boldsymbol{\Theta}$.

2.3.2.1 Independence Models. The simplest non dynamic model is the so called *independence model*.

Definition 2.3.4. *An* independence model *for $(\boldsymbol{Y}, \boldsymbol{\theta})$ is such that $\perp\!\!\!\perp_{i \in [n]} Y_i \mid \boldsymbol{\theta}$.*

The independence model class can be simply described by a graph with vertex set $\{Y_i : i \in [n]\}$ and empty edge set.

2.3.2.2 Bayesian Networks. BNs [Jensen and Nielsen, 2009, Korb and Nicholson, 2003, Lauritzen, 1996, Pearl, 1988, Smith, 2010] are the graphical model most widely used in practice. There are now thousands of applications of these probabilistic tools in a variety of domains [Aguilera et al., 2011, Cowell et al., 2011, Heckerman et al., 1995, Niedermayer, 2008, Uusitalo, 2007]. These are especially attractive because the Directed Acyclic Graph (DAG) associated to a BN embodies qualitative information in terms of various conditional independences, and so often easier to elicit and explain then some of its competitors. Explicitly, in a BN each random variable of the problem corresponds to a vertex of the DAG and directed edges between vertices represent possible dependencies between the variables.

Only density functions associated to certain sets of conditional independences can be represented by a specific DAG model. The following definition introduces one such class of conditional independence requirements. We let $A_i' \subset [n]$ be the index set of the non descendants of Y_i, whilst $\Pi_i \subset [n]$ is the index set of its parents, $i \in [n]$.

Definition 2.3.5. *A probability density f over $\boldsymbol{Y} \mid \boldsymbol{\theta}$ is said to obey the* local directed Markov property *relative to a DAG $\mathcal{G}$, if, for $i \in [n]$, $Y_i \perp\!\!\!\perp \boldsymbol{Y}_{A_i'} \mid \boldsymbol{Y}_{\Pi_i}, \boldsymbol{\theta}$.*

Therefore, a DAG is able to describe the conditional independences of a random vector only if the associated density is local directed Markov. Cowell et al. [1999], among the others, showed that there are other equivalent statements to the local directed Markov property that can characterise probability densities respecting the topology of a DAG. One such condition is the so called recursive factorisation property of the density function.

Lemma 2.3.6. *If a density f over $\boldsymbol{Y}$ obeys the local directed Markov property relative to $\mathcal{G}$, then it respects the* recursive factorisation *formula, i.e.*

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = \prod_{i \in [n]} f(y_i \mid \boldsymbol{y}_{\Pi_i}, \boldsymbol{\theta}_i)$$

The BN model can now be formally defined as follows.

Definition 2.3.7. *A BN over $\boldsymbol{Y}$ consists of*

- *a DAG $\mathcal{G}$ with vertex set $\{Y_i : i \in [n]\}$ and an edge from Y_i to Y_j if and only if (iff) $i \in \Pi_j$;*

- *for $i \in [n]_1 = [n] \setminus \{1\}$, conditional independence statements of the form $Y_i \perp\!\!\!\perp \boldsymbol{Y}_{A_i'} \mid \boldsymbol{Y}_{\Pi_i}, \boldsymbol{\theta}$;*

- *for $i \in [n]$, conditional densities $f(y_i \mid \boldsymbol{y}_{\Pi_i}, \boldsymbol{\theta}_i)$.*
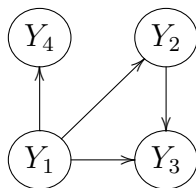
FIGURE 2.1: Example of a directed acyclic graph depicting the relationships between four random variables.

As briefly pointed out previously, one of the main strengths of BN modelling lies in its qualitative nature. Relationships between random variables are intuitively depicted by a graph, which are therefore easy to justify and explain in natural language. Smith [2010] described a protocol that can be followed to elicit the graph structure associated to a BN, whilst O'Hagan [2006] discussed how the associated densities can be quantified. Although in this thesis we always assume that the structure of the DAG of a BN is elicited by either a DM or appropriate experts, we note here that several routines to automatically learn the DAG's topology from data are now in place under a variety of conditions [see e.g. Cooper and Yoo, 1999, Neapolitan, 2004].

Example 2.3.8. Consider the DAG in Figure 2.1 with vertex set $\{Y_1, Y_2, Y_3, Y_4\}$ and suppose

- Y_1: amount of contamination;
- Y_2: human radioactive intake;
- Y_3: effects on human health;
- Y_4: political disruption.

The only conditional independence statement associated to the DAG of Figure 2.1 is $Y_4 \perp\!\!\!\perp Y_3, Y_2 \mid Y_1, \boldsymbol{\theta}$, implying that, once the amount of contamination is known, human intake and effects on health are irrelevant to predict the political disruption. The recursive factorisation associated to this DAG corresponds to

$$f(y_1, y_2, y_3, y_4 \mid \boldsymbol{\theta}) = f(y_4 \mid y_1, \boldsymbol{\theta}_4) f(y_3 \mid y_2, y_1, \boldsymbol{\theta}_3) f(y_2 \mid y_1, \boldsymbol{\theta}_2) f(y_1 \mid \boldsymbol{\theta}_1).$$

The $n - 1$ conditional independence statements defining a BN model are not the only ones implied by a given model. The **d-separation** criterion [Lauritzen, 1996] provides a rule to verify whether or not a generic conditional independence statement holds for a BN model. For three disjoint subsets A, B and C of $[n]$, the conditional independence $\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_C, \boldsymbol{\theta}$ holds iff $\{Y_i : i \in C\}$ separates $\{Y_i : i \in A\}$ from $\{Y_i : i \in B\}$ in the moral graph of $\mathcal{G}$ (see Appendix B). Importantly, a set of conditional independence statements does not uniquely identify a DAG and several DAGs can represent the same set of conditional independences. Such DAGs are said to be *equivalent*. So for example, the DAG in Figure 2.1 would still imply the same conditional independence structure if

the edge between Y_2 and Y_3 were reversed. We discuss in more details in Section 2.3.4 how this issue is related to the concept of statistical causality.

We further note that the BN model provides the basis for fast computations of probabilities through propagation techniques over the graph. Propagation is usually performed over the junction tree of the DAG [see e.g. Smith, 2010], but also other, often approximated, methodologies exist [Cowell et al., 1999, Murphy et al., 1999]. Spiegelhalter and Lauritzen [1990] introduced two conditions over the parameter vector $\boldsymbol{\theta}$ that can allow for computations to be fast but still exact.

Definition 2.3.9. *For a BN over $\boldsymbol{Y}$ we say that $\boldsymbol{\theta}$ respects the **global independence** condition if $\perp\!\!\!\perp_{i \in [n]} \boldsymbol{\theta}_i$.*

Definition 2.3.10. *Let $\boldsymbol{\theta}_i = (\theta_{ji})^{\mathrm{T}}_{j \in [n_i]}$, $n_i \in \mathbb{Z}_{\geq 1}$, $i \in [n]$. We say that the **local independence** condition holds for a BN over $\boldsymbol{Y}$ if, for every $i \in [n]$, $\perp\!\!\!\perp_{j \in [n_i]} \theta_{ji}$.*

Together, local and global independence imply that the parameters are mutually independent of each other. In Chapter 3 we introduce a generalisation of global independence for BN models whose densities are elicited separately by different groups of experts and show that this condition entails distributed group inferences.

Note that the notion of global independence cannot hold unless the parameter spaces are *variationally independent* [Dawid, 2001, Dawid and Lauritzen, 1993].

Definition 2.3.11. *If $\boldsymbol{\Theta} = \bigtimes_{i \in [n]} \boldsymbol{\Theta}_i$ we say that the spaces $\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_n$ are variationally independent.*

Under the global independence assumption the following result from Spiegelhalter and Lauritzen [1990] hold.

Proposition 2.3.12. *If global independence holds for a BN with vertex set $\{Y_i : i \in [n]\}$, then*

$$f(\boldsymbol{y}) = \int_{\boldsymbol{\Theta}} f(\boldsymbol{y}, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} = \int_{\boldsymbol{\Theta}} \prod_{i \in [n]} f(y_i \mid \boldsymbol{y}_{\Pi_i}, \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i) \mathrm{d}\boldsymbol{\theta} = \prod_{i \in [n]} f(y_i \mid \boldsymbol{y}_{\Pi_i}), \qquad (2.3.1)$$

where

$$f(y_i \mid \boldsymbol{y}_{\Pi_i}) = \int_{\boldsymbol{\Theta}_i} f(y_i \mid \boldsymbol{y}_{\Pi_i}, \boldsymbol{\theta}_i) \pi(\boldsymbol{\theta}_i) \mathrm{d}\boldsymbol{\theta}_i. \qquad (2.3.2)$$

Therefore under the assumption of global independence also the marginal distribution over $\boldsymbol{Y}$ factorises according to the underlying DAG. This would not be true in general without the global independence assumption.

Another useful property arising from global independence is that Bayesian updating given a sample $\boldsymbol{x}$ can be performed in a distributed fashion. The sample $\boldsymbol{x}$ does not need to be complete in order for this to be the case as the following proposition from Spiegelhalter and Lauritzen [1990] formalises. We let $Fa_i \subseteq [n]$ and $De_i \subset [n]$ be the index sets of the family and the descendants of Y_i, respectively, $i \in [n]$.

Proposition 2.3.13. *Suppose $\boldsymbol{x}$ is an ancestral sample from the same population of $\boldsymbol{Y}$, meaning that if $\boldsymbol{x}_i$, the sample relative to Y_i, is observed, then $\boldsymbol{x}_{A_i'}$ is observed as well. Then, if global independence holds and letting I be the set of indices of either the vertices whose children are not sampled or leaves of the graph, we have that*

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) = \prod_{i \in I} \prod_{j \in A_i} \pi(\boldsymbol{\theta}_j \mid \boldsymbol{x}_{Fa_j}) \prod_{k \in De_i} \pi(\boldsymbol{\theta}_k). \tag{2.3.3}$$

Similar recursions hold for local independence as well which retain the independence of the parameters a posteriori [see Smith, 2010, Spiegelhalter and Lauritzen, 1990].

We next consider Gaussian BNs, which define a Gaussian distribution for the random vector $\boldsymbol{Y}$ respecting the conditional independences implied by a DAG. Given a DAG $\mathcal{G}$, we first define each vertex as a simple linear regression.

Definition 2.3.14. *Let $\mathcal{G}$ be the DAG of a BN with vertex set $\{Y_i : i \in [n]\}$. We say that a vertex Y_i, $i \in [n]$, is defined as a* DAG *linear regression if*

$$Y_i = \theta_{0i} + \sum_{j \in \Pi_i} \theta_{ji} Y_j + \varepsilon_i, \tag{2.3.4}$$

where $\theta_{ji} \in \mathbb{R}$ are regression parameters and the error $\varepsilon_i \sim (0, \psi_i)$, i.e. ε_i has mean zero and variance $\psi_i \in \mathbb{R}_{>0}$.

We extensively deal with these types of models in Chapter 5.

The following proposition, from Richardson and Spirtes [2002] and expressed here in the form of Sullivant et al. [2010], then shows how to build a multivariate Normal distribution from these regression specifications respecting the conditional independence statements associated to the graph.

Proposition 2.3.15. *Let $\boldsymbol{Y} = (Y_i)_{i \in [n]}^{\mathrm{T}}$ be a random vector, where each Y_i is defined as a DAG linear regression, and $\boldsymbol{\varepsilon}_i = (\varepsilon_i)_{i \in [n]}^{\mathrm{T}}$. Assume that $\perp\!\!\!\perp_{i \in [n]} \varepsilon_i$, $\boldsymbol{\varepsilon} \perp\!\!\!\perp \boldsymbol{Y}$ and $\varepsilon_i \sim \mathrm{N}(0, \psi)$, i.e. the errors follow a Normal distribution. Let $\Psi = \mathrm{diag}(\psi_1, \dots, \psi_n)$, meaning that Ψ is a diagonal matrix, and L be an upper triangular $n \times n$ matrix defined as*

$$L_{ji} = \begin{cases} \theta_{ji}, & \text{if } j \in \Pi_i, \\ 0, & \text{otherwise.} \end{cases}$$
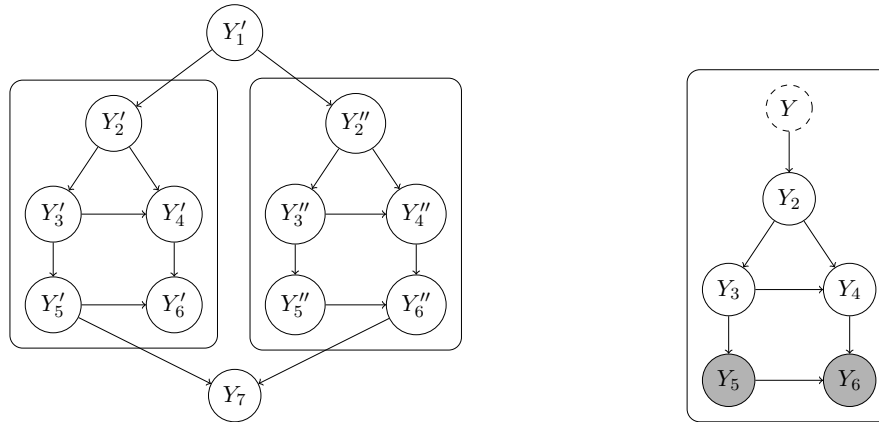
(A) Example of a Bayesian network with identical subnetworks. (B) Class definition for the sub-networks of Figure 2.2a.

FIGURE 2.2: Example of object orientation in Bayesian networks.

Define $B = I_n - L$, where I_n is the $n \times n$ identity matrix, and $\boldsymbol{\theta}_0 = (\theta_{0i})_{i \in [n]}^{\mathrm{T}}$. Then
$$\boldsymbol{Y} \sim \mathrm{N}\left((B^{-1})^{\mathrm{T}}\boldsymbol{\theta}_0, (B^{-1})^{\mathrm{T}}\Psi B^{-1}\right).$$

In a full Bayesian framework prior distributions for the regression parameters θ_{ji} and the error variances ψ_i need to be further provided, as, for example, the Normal Inverse Gamma for the simple linear model shown in Appendix C.2.1.

To illustrate global and local independence for Gaussian BNs, let $\boldsymbol{\theta}_i^{\mathrm{T}} = (\theta_{0i}, \psi_i, \theta_{ji})_{j \in \Pi_i}$, $i \in [n]$. Global independence is then such that $\perp\!\!\!\perp_{i \in [n]} \boldsymbol{\theta}_i$. Local independence on the other hand assumes that every element of $\boldsymbol{\theta}_i$ is independent of the rest, for every $i \in [n]$.

2.3.2.3 Object Oriented Bayesian Networks. In complex large BNs identical subnetworks are often repeated in different parts of the DAG. This is for example the case in the DAG of Figure 2.2a, where the subnetworks within the two rectangles have the same topology. Now assuming that Y_i' and Y_i'', $i \in [6]_1$, have the same sample space and equal conditional probability densities, we can then represent the network as an Object Oriented Bayesian Network (OOBN). As extensively discussed in Chapter 1, the object oriented paradigm has proven to provide a convenient framework to represent complex problems. We do not formally define this class of models [see Koller and Pfeffer, 1997], since this is beyond the scope of the thesis and requires the introduction of a rather intricate notation. However, following Jensen et al. [2006] and Neil et al. [2000], we still provide an intuitive example of this fairly recent technology.

Considering again the network in Figure 2.2a, each subnetwork within the rectangles in the object oriented paradigm is an instantiation of a *class*, called *object*. Figure
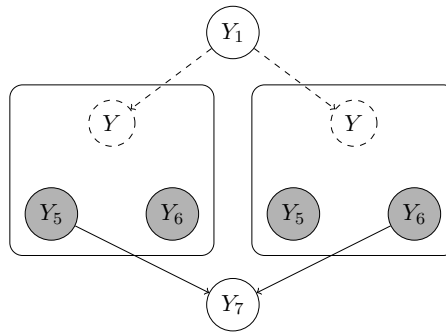
FIGURE 2.3: Object oriented representation of the Bayesian network of Figure 2.2a, whose class is defined in Figure 2.2b.

2.2b shows the class definition for the repeated subnetworks of Figure 2.2a. Each class consists of three types of vertices:

- input vertices: not true random variables, but artificial nodes that serve as 'placeholders'. In Figure 2.2b the input vertex is the dashed node Y which must have the same state space of Y_1;

- encapsulated vertices: vertices that are hidden outside of the class and that are used only within the class. In Figure 2.2b the encapsulated vertices are Y_2, Y_3 and Y_4;

- output vertices: part of the class that is accessible from outside and that can be connected to other classes. The output vertices of the class in Figure 2.2b are the shaded nodes Y_5 and Y_6.

Encapsulated and output vertices are allowed to be objects, whilst input vertices must be variables.

Given the class definition in Figure 2.2b, we can represent the BN of Figure 2.2a as the OOBN reported in Figure 2.3. In this network the encapsulated vertices are hidden and only input and output nodes of the class explicitly appear. Each rectangle is an object and in each object the relative output nodes are connected to elements external to the class. The dashed arrows into the input node indicate which node Y is a placeholder for (in this case Y_1 for both objects).

Even in this simple example the expressive power of OOBNs becomes apparent, modularising the overall problem in different layers formally separated. Of course in much larger examples OOBNs provide an even more compact representation of the domain under consideration: see for example Neil et al. [2000].

2.3.2.4 Probabilistic Undirected Graphs. We now consider probability densities f that can be described by an Undirected Graph (UG) $\mathcal{G}$. These are used in particular when there is no clear directional relationships between the variables under study. For the purpose of the thesis we assume $\mathcal{G}$ to be always decomposable in the undirected case: we motivate this assumption below. Just as in the directed case of Section 2.3.2.2, we first introduce a class of densities which are in some sense compatible to the graph and then define the class of UG models called *Markov Networks (MNs)*. References about these models are countless [see e.g. Castillo, 1997, Cowell et al., 1999, Lauritzen, 1996]. We follow here Dawid and Lauritzen [1993].

Definition 2.3.16. *Let* $A, B \subset [n]$ *and* $V(\mathcal{G}) = \{Y_i : i \in [n]\}$ *be the vertex set of an UG* $\mathcal{G}$. *A density* f *over* $\boldsymbol{Y} \mid \boldsymbol{\theta}$ *is called* Markov *if, for any decomposition* $\{Y_i : i \in A\}$, $\{Y_i : i \in B\}$ *of* $\mathcal{G}$, $\boldsymbol{Y}_A \perp\!\!\!\perp \boldsymbol{Y}_B \mid \boldsymbol{Y}_{A\cap B}, \boldsymbol{\theta}$

Definition 2.3.17. *A decomposable MN model consists of a decomposable UG* $\mathcal{G}$ *with vertex set* $\{Y_i : i \in [n]\}$ *together with a Markov density* f.

Note that for each vertex Y_i of an MN, the conditional independence statement

$$Y_i \perp\!\!\!\perp \boldsymbol{Y}_{[n]\setminus\{i\}} \mid \boldsymbol{Y}_{Ne_i}, \boldsymbol{\theta},$$

holds, where Ne_i is the index set of the neighbours of Y_i. Thus, just as in the directed case, Markov densities associated to a (decomposable) UG are the only ones entertaining the separations implied by the graph. In order to deduce the associated Markov factorisation as in the directed case, we first need to introduce the notion of *consistency*.

Definition 2.3.18. *Let* f_A *and* f_B *be densities over* $\boldsymbol{Y}_A \mid \boldsymbol{\theta}$ *and* $\boldsymbol{Y}_B \mid \boldsymbol{\theta}$ *respectively,* $A, B \in [n]$. *We say that* f_A *and* f_B *are consistent if they both yield the same density over* $\boldsymbol{Y}_{A\cap B} \mid \boldsymbol{\theta}$.

Now letting $\mathcal{C}$ and $\mathcal{S}$ be the sets of the indices of the cliques and the separators respectively of an UG $\mathcal{G}$, we have the following.

Lemma 2.3.19. *The unique Markov density* f *over a decomposable UG* $\mathcal{G}$ *having consistent densities as its clique marginals factorises as*

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = \frac{\prod_{C \in \mathcal{C}} f(\boldsymbol{y}_C \mid \boldsymbol{\theta}_C)}{\prod_{S \in \mathcal{S}} f(\boldsymbol{y}_S \mid \boldsymbol{\theta}_S)^{v_S}},$$

where v_S *is the multiplicity of the separator* S, $\boldsymbol{\theta}_C$, $\boldsymbol{\theta}_S$ *are the parameter vectors associated to* $\boldsymbol{Y}_C$ *and* $\boldsymbol{Y}_S$, *respectively, and* $\mathcal{C}$ *and* $\mathcal{S}$ *are the index sets of the cliques and the separators, respectively.*
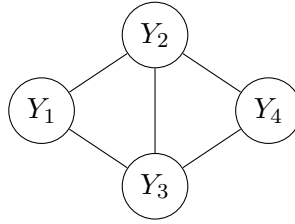
FIGURE 2.4: Example of a decomposable undirected graph with four vertices.

Given an order over the elements of $\mathcal{C}$ exhibiting the running intersection property (see Appendix B), the density of a Markov density over a decomposable graph can be alternatively written as, assuming the graph includes m cliques,

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = \prod_{i\in[m]_1} f(\boldsymbol{y}_{R_i} \mid \boldsymbol{y}_{S_i}, \boldsymbol{\theta}_{R_i}) f(\boldsymbol{y}_{C_1} \mid \boldsymbol{\theta}_{C_1}),$$

where $R_i = C_i \setminus S_i$.

Example 2.3.20. Consider the UG with vertex set $\{Y_1, Y_2, Y_3, Y_4\}$ in Figure 2.4. The conditional independence statement associated to this graph is $Y_4 \perp\!\!\!\perp Y_1 \mid Y_2, Y_3, \boldsymbol{\theta}$, and therefore the associated density factorises as

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = \frac{f(y_1, y_2, y_3 \mid \boldsymbol{\theta}_{[3]}) f(y_2, y_3, y_4 \mid \boldsymbol{\theta}_{[4]_1})}{f(y_2, y_3 \mid \boldsymbol{\theta}_{[3]_1})},$$

or alternatively as

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = f(y_4 \mid y_2, y_3, \boldsymbol{\theta}_4) f(y_1, y_2, y_3 \mid \boldsymbol{\theta}_{[3]}) = f(y_1 \mid y_2, y_3, \boldsymbol{\theta}_1) f(y_2, y_3, y_4 \mid \boldsymbol{\theta}_{[4]_1}).$$

Just as in the directed case, conditions over the parameter vector can be imposed entailing distributed inferences. For the purpose of this thesis we introduce only the strong hyper Markov condition of Dawid and Lauritzen [1993].

Definition 2.3.21. *A density π is strong hyper Markov for $\mathcal{G}$ if, for any decomposition $\{Y_i : i \in A\}$, $\{Y_i : i \in B\}$ of $\mathcal{G}$, $\boldsymbol{\theta}_A \perp\!\!\!\perp \boldsymbol{\theta}_B$.*

Since the cliques of a decomposable UG can be ordered to sequentially form a decomposition, we have the following result from Dawid and Lauritzen [1993].

Lemma 2.3.22. *If a density π for $\boldsymbol{\theta}$ is strong hyper Markov then $\pi(\boldsymbol{\theta}) = \prod_{C\in\mathcal{C}} \pi(\boldsymbol{\theta}_C)$.*

We do not show in detail how to build a distribution exhibiting the strong hyper Markov condition. However we briefly note here that this can be straightforwardly done by considering the Markov combinations of the marginal distributions over the cliques of the graph [Massa and Lauritzen, 2010].

The following proposition from Dawid and Lauritzen [1993] then shows how the strong hyper Markov condition is associated to fast distributed Bayesian inference.

Proposition 2.3.23. *Assume a vector $\boldsymbol{X}$ is sampled from the same population of $\boldsymbol{Y}$ and assume the density π of $\boldsymbol{\theta}$ is strong hyper Markov. Then, the posterior distribution obtained by conditioning on the complete data $\boldsymbol{X} = \boldsymbol{x}$ is the unique strong hyper Markov distribution specified by the clique-marginal prior distributions given by $\pi(\boldsymbol{\theta}_C \mid \boldsymbol{x}) = f(\boldsymbol{x}_C \mid \boldsymbol{\theta}_C)\pi(\boldsymbol{\theta}_C)$.*

Therefore Bayesian updating can be performed locally within each clique whilst retaining the strong hyper Markov condition. Note that the above result does not hold in general for non-decomposable undirected models. Whilst global independence is retained after observing certain incomplete datasets, strong hyper Markov laws are also broken by any incomplete observation. In Chapter 3 however we are able to show that for certain sets of incomplete observation the strong hyper Markov property can be retained if Bayesian updating is performed simultaneously for all the elements of this set.

We now consider the class of MN models whose associated distribution is Gaussian. A *Gaussian MN* with decomposable UG $\mathcal{G}$, $V(\mathcal{G}) = \{Y_i : i \in [n]\}$, is defined as $\boldsymbol{Y} \mid \boldsymbol{\mu}, \Sigma \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (\mu_i)_{i \in [n]}^{\mathrm{T}} \in \mathbb{R}^n$ and Σ is an $n \times n$ covariance matrix[3] such that if $(Y_i, Y_j) \notin E(\mathcal{G}) \implies \Sigma_{ij}^{-1} = 0$, where Σ_{ij}^{-1} is the entry of Σ^{-1} in position (i, j). So for example the covariance matrix associated to the Gaussian MN with graph in Figure 2.4 is such that its inverse has zero entries in positions $(1, 4)$ and $(4, 1)$. Now for simplicity let $\boldsymbol{\mu} = \boldsymbol{0}$. Such a model class is usually referred to as a covariance selection model [Wermuth, 1976]. Recall from Appendix C.2.3 that the Inverse Wishart distribution allows for conjugate learning with Normal models when defined as a prior of the covariance matrix. Let $\Sigma \sim \mathrm{IW}(A, d)$, where A is a positive semidefinite $n \times n$ matrix and $d \in \mathbb{Z}_{\geq n}$. It is well known that for a subset $B \in [n]$, $\boldsymbol{Y}_B \sim \mathrm{N}(\boldsymbol{0}, \Sigma_{B,B})$ and $\Sigma_{B,B} \sim \mathrm{IW}(A_{B,B}, d)$, where, for a matrix Σ, $\Sigma_{B,B}$ is its submatrix with rows and columns $i \in B$ [see e.g. Dawid and Lauritzen, 1993]. For every clique $C \in \mathcal{C}$ of $\mathcal{G}$, let $\boldsymbol{Y}_C \mid \Sigma_{C,C} \sim \mathrm{N}(\boldsymbol{0}, \Sigma_{C,C})$ and assume the covariance $\Sigma_{C,C}$ has an Inverse Wishart prior distribution $\mathrm{IW}(A^C, d)$. A unique strong hyper Markov distribution for the whole graph $\mathcal{G}$ having as marginals over the cliques these Inverse Wishart distributions exists if, calling $S_{ij} = C_i \cap C_j$, the matrices $A_{S_{ij}, S_{ij}}^{C_i}$ and $A_{S_{ij}, S_{ij}}^{C_j}$ are identical.[4] We then say that Σ follows an hyper Inverse Wishart distribution with parameters A and d, $\Sigma \sim \mathrm{HIW}(A, d)$, where A is such that $A_{C,C} = A^C$, for $C \in \mathcal{C}$.

---

[3] A covariance matrix is a symmetric positive semidefinite matrix with entries in $\mathbb{R}_{>0}$ on the diagonal and in $\mathbb{R}$ otherwise.

[4] This result follows from Proposition 5.9 of Dawid and Lauritzen [1993] which we have not introduced here since it would require the introduction of additional concepts not relevant for the thesis. In a nutshell, this result is true because the Inverse Wishart is conjugate for the Normal covariance selection model and this distribution is a member of the exponential family.

For the graph in Figure 2.4, let $\Sigma_{[3],[3]} \sim \mathrm{IW}(A^{123}, d)$ and $\Sigma_{[4]_1,[4]_1} \sim \mathrm{IW}(A^{234}, d)$. Then an hyper Inverse Wishart distribution for this graph having this marginals exists if $A^{234}_{[3]_1,[3]_1} = A^{123}_{[3]_1,[3]_1}$.

Suppose now that a vector of observations $\boldsymbol{x} = (\boldsymbol{x}_i^{\mathrm{T}})_{i\in[m]}^{\mathrm{T}}$ from the same family of $\boldsymbol{Y}$ has been collected, where $\boldsymbol{x}_i = (x_{ij})_{j\in[n]}^{\mathrm{T}}$. Assume an hyper Markov law is constructed from the clique marginals as exemplified above. Then from Appendix C.2.3 and Proposition 2.3.23 we know that for the covariance selection model the posterior density of $\Sigma \mid \boldsymbol{x}$ is strong hyper Markov, where each marginal over the submatrix associated with a clique $\Sigma_{C,C}$, $C \in \mathcal{C}$, follows an Inverse Wishart with parameters $A^C + mS_{C,C}$ and $d + m$, where S is equal to $m^{-1} \sum_{i\in[m]} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}$ (see Appendix C.2.3). This can be easily noted within our example. The marginal posteriors over the cliques are $A^{123} \sim \mathrm{IW}(A^{123} + mS_{[3],[3]}, d+m)$ and $A^{234} \sim \mathrm{IW}(A^{234} + mS_{[4]_1,[4]_1}, d + m)$. It then holds that $A^{123}_{[3]_1,[3]_1} + mS_{[3]_1,[3]_1} = A^{234}_{[3]_1,[3]_1} + mS_{[3]_1,[3]_1}$, where

$$mS_{[3]_1,[3]_1} = \begin{pmatrix} \sum_{i\in[m]} x_{i2}^2 & \sum_{i\in[m]} x_{i2}x_{i3} \\ \sum_{i\in[m]} x_{i2}x_{i3} & \sum_{i\in[m]} x_{i3}^2 \end{pmatrix} \tag{2.3.5}$$

2.3.2.5   Probabilistic Chain Graphs. Probabilistic Chain Graph (PCG) models are a hybrid representation of undirected and directed graphical models, allowing for the underlying graph $\mathcal{G}$ to be mixed and more specifically to be a Chain Graph (CG). We give here only a brief introduction to this class of models [see e.g Cowell et al., 1999, Drton, 2009, Frydenberg, 1990, Lauritzen, 1996, for more details]. These are especially useful when there are both directional and non-directional relationships between the variables associated to the vertices of the underlying graph. As with the earlier models, we first introduce the relative Markov property, then derive the associated factorisation and in conclusion define the model. We let Bd_i be the index set of the variables in the boundary of Y_i, $i \in [n]$.

Definition 2.3.24. *A probability density f over $\boldsymbol{Y} \mid \boldsymbol{\theta}$ is said to obey the chain Markov property relative to a CG $\mathcal{G}$ with vertex set $\{Y_i : i \in [n]\}$ and m strong components if, for every vertex Y_i, $i \in [n]$, $Y_i \perp\!\!\!\perp \boldsymbol{Y}_{A_i' \setminus Bd_i} \mid \boldsymbol{Y}_{Bd_i}, \boldsymbol{\theta}$.*

Proposition 2.3.25. *If a density f obeys the chain Markov property relative to a CG $\mathcal{G}$ with vertex set $\{Y_i : i \in [n]\}$, then its density factorises as*

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = \prod_{i\in[m]} f\left(\boldsymbol{y}_{C_i} \mid \boldsymbol{y}_{\Pi_{C_i}}, \boldsymbol{\theta}_{C_i}\right),$$

where $C_1, \ldots, C_m$ are the index sets of the variables in the strong components of $\mathcal{G}$ and $\Pi_{C_i} = \cup_{j\in C_i} \Pi_j$.
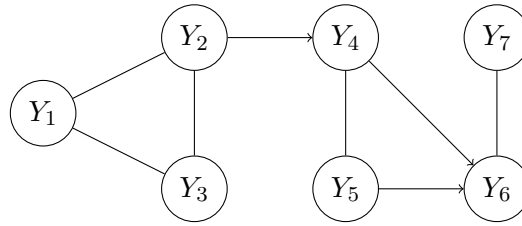
FIGURE 2.5: Example of a probabilistic chain graph model.
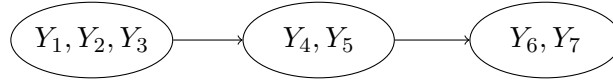
FIGURE 2.6: Representation of the chain graph in Figure 2.5 as a directed acyclic graph.

We are now ready to formally define the model.

Definition 2.3.26. *A PCG model consists of a CG $\mathcal{G}$ together with a density f respecting the chain Markov property.*

Example 2.3.27. Consider the PCG in Figure 2.5. It can be deduced from this graph that the following conditional independence statements need to hold for a chain Markov distribution

$$Y_4 \perp\!\!\!\perp Y_1, Y_3 \mid Y_2, Y_5, \boldsymbol{\theta}, \qquad Y_5 \perp\!\!\!\perp Y_1, Y_2, Y_3 \mid Y_4, \boldsymbol{\theta},$$
$$Y_6 \perp\!\!\!\perp Y_1, Y_2, Y_3 \mid Y_4, Y_5, Y_7, \boldsymbol{\theta}, \qquad Y_7 \perp\!\!\!\perp Y_1, \ldots, Y_5 \mid Y_6, \boldsymbol{\theta},$$

and the associated density can be factored as

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = f(y_6, y_7 \mid y_4, y_5, \boldsymbol{\theta}_{67}) f(y_4, y_5 \mid y_3, \boldsymbol{\theta}_{45}) f(y_1, y_2, y_3 \mid \boldsymbol{\theta}_{123}).$$

Note that each PCG can be represented by a less expressive DAG whose vertex set corresponds to the strong components of the underlying CG. So for example the PCG in Figure 2.5 can be transformed into the DAG of Figure 2.6.

2.3.2.6 Staged and Event Trees. All the classes of models defined so far are able to depict standard conditional independence statements only. The class of *event tree* models we discuss in this section is able to explicitly represent context specific independences. Event trees are such that their nodes are the situations in which a process might find itself and the edges emanating from a node are the possible unfoldings given the current situation. It has been extensively discussed in the literature that these models are extremely expressive in describing how processes might unfold. These are especially useful in the cases where the variables are ordered in a way that follows the narrative of the events [Freeman and Smith, 2011, Shafer, 1996, Smith, 2010].
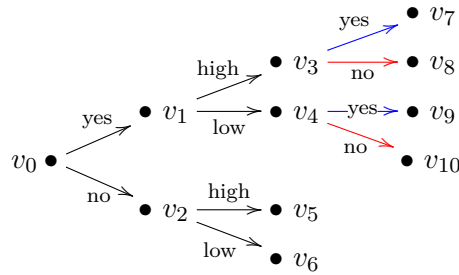
FIGURE 2.7: Example of an event tree on three variables with the additional representation of its stages in different colors.

Let $\mathcal{T}$ be a directed tree, $\Lambda(v, \mathcal{T})$ denote the set of paths from $v \in V(\mathcal{T})$ to a leaf node of $\mathcal{T}$ and $\mathcal{Y} = \Lambda(s_0, \mathcal{T})$, where s_0 is the root of the tree, be the set of root-to-leaf paths. Each path $y \in \mathcal{Y}$ is a so-called atomic event, i.e. a possible unfolding of events. Finally let $\mathcal{Y}_s$ denote the set of children of a situation $s \in S(\mathcal{T})$.

Definition 2.3.28. *An **event tree** is a directed tree $\mathcal{T}$ together with a random variable Y_s for each situation $s \in S(\mathcal{T})$ with sample space $\mathcal{Y}_s$ defined conditional on having reached vertex s. The distribution of Y_s is determined by the* floret probability vector $\boldsymbol{\theta}_s = (\theta_{ss'})_{s' \in \mathcal{Y}_s}^{\mathrm{T}}$, *where $\theta_{ss'} = \mathbb{P}(Y_s = s')$.*

The expressive power of event trees can be increased by identifying probabilities associated to different edges that are equal. Trees can then be embellished by a colouring of the edges, where two edges have same colour if their associated probabilities are equal.

Definition 2.3.29. *A **staged tree** is an event tree where, for some $s, s' \in S(\mathcal{T})$, the floret probability vectors are identified, $\boldsymbol{\theta}_s = \boldsymbol{\theta}_{s'}$, and equally coloured. Then s and s' are in the same* stage. *We let $\mathbb{W}$ be the set of stages of $\mathcal{T}$.*

Example 2.3.30. Suppose a problem is modelled with three binary random variables: Y_1, release from a source term; Y_2, level of contamination in the surrounding area; Y_3, political disruption in the region due to the release. Suppose $\mathcal{Y}_1 = \mathcal{Y}_3 = \{\text{yes}, \text{no}\}$ and $\mathcal{Y}_2 = \{\text{high}, \text{low}\}$. It is assumed that, if there has been a release, the level of contamination does not provide any information to predict the political disruption. This problem could then be modelled by a BN with vertex set $\{Y_1, Y_2, Y_3\}$ and edge set $\{(Y_1, Y_2), (Y_1, Y_3)\}$.

However the BN representation forces us to retain information which is meaningless in this context, as for instance the atom (no,high,yes) which has probability zero. This is on the other hand explicitly modelled in the staged tree in Figure 2.7. In this tree the leftmost two edges are the possible outcomes of Y_1, the edges in the center are the outcomes of Y_2 given the different levels of Y_1, whilst the rightmost edges coincide with the outcomes of Y_3 given Y_2 and Y_1. In the lower part of the tree, associated
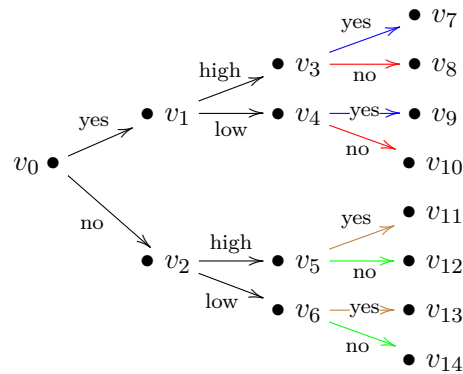
FIGURE 2.8: Tree representation of the Bayesian network of Example 2.3.30 .

to $Y_1 = $ no, there are no outcomes for the political disruption variable, as these have probability zero. The set of situations of this tree includes vertices $\{v_0, \ldots, v_6\}$ and leaves $\{v_7, \ldots, v_{10}\}$. Its stages are $w_0 = \{v_0\}$, $w_1 = \{v_1\}$, $w_2 = \{v_2\}$, $w_3 = \{v_3, v_4\}$. Furthermore $\mathbb{W} = \{w_0, \ldots, w_3\}$. The tree representation of the associated BN model is in Figure 2.8. Such a tree is extremely regular in its colouring and each of its root to leaf paths have the same length: a property exhibited by the tree representation of any BN (see for more details Section 5.6.2).

Although the colouring of the edges entails an improved representation of the overall structure of the problem at hand, conditional independences are still difficult to read from the graph. In addition as the number of variables increases, the size of the tree becomes quickly too large to be concisely reported. The class of models we introduce in the following section is able to compactly represent every conditional independence entertained.

2.3.2.7 Chain Event Graphs. Chain Event Graphs (CEGs) [Smith and Anderson, 2008] are models capable of representing context specific conditional independences in a single compact graphical representation. CEGs are constructed by starting with a staged tree, therefore requiring variables to be discrete, and then merging into a single vertex certain situations that are in the same stage.

Definition 2.3.31. *We say that two situations s and s' are in the same* position *if the subtrees with roots s and s', respectively, have the same topology and the same edge colouring. We let $\mathbb{B}$ denote the set of positions of a staged tree.*

Note that two situations in the same position are by definition also in the same stage, whilst it does not necessary follows that two situations in the same stage are also in the same position.
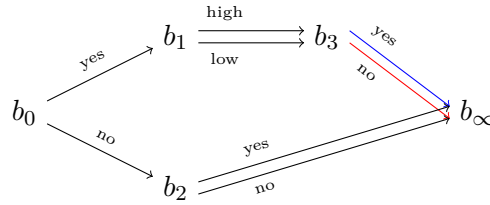
FIGURE 2.9: A chain event graph model representing the staged tree in Figure 2.7.

Example 2.3.32. Considering the staged tree in Figure 2.7, we can see that for this example the set of stages coincides with the one of positions.

Definition 2.3.33. *A **CEG** is the graph obtained by collapsing a staged tree into its positions. Its vertex set is equal to $\{\mathbb{B} \cup b_\infty\}$, where b_∞ is the vertex collecting all the leaves of the tree. Its edge set is such that*

- *there is an edge (b_i, b_j), $b_i, b_j \in \mathbb{B}$, for every edge from a generic vertex $v \in b_i$ to any vertex $v' \in b_j$;*

- *there is an undirected edge between any two positions in the same stage;*

Example 2.3.34. The CEG representation of the staged tree of Figure 2.7 is shown in Figure 2.9, where we further annotated and coloured the edges as in Barclay et al. [2013].

We chose the CEG as a representative model for context specific independences for two main reasons. First, compared to other models capable of representing context specific conditional independences, the CEG consists of a single graph. Furthermore Smith and Anderson [2008] showed that every discrete BN can be represented as a CEG, whilst the converse does not hold. This is not case for example for probabilistic decision graphs [Jaeger, 2004] and context specific BNs [Boutilier et al., 1996]. Barclay et al. [2013] nicely discussed how to convert a BN into a CEG model and how to measure the advantages of the latter representation. Second, a wide range of methods to perform various statistical analyses have been developed for this model class [see e.g. Barclay et al., 2013, 2014, Freeman and Smith, 2011, Thwaites and Smith, 2015, Thwaites et al., 2010].

Importantly, under assumptions similar to local and global independence, but customised to staged trees and CEGs, Bayesian updating can be performed in a distributed way through the Multinomial-Dirichlet recursions illustrated in Appendix C.1.2. In Chapter 3 we generalise these independences for CEGs to take into account that probabilities over the graph are elicited by different groups of experts. Suppose $w_i \in \mathbb{W}$ has n_i emanating edges with associated probability vector $\boldsymbol{\theta}_i = (\theta_{ij})_{j \in [n_i]}^{\mathrm{T}}$, such that $\theta_{ij} \in (0, 1)$ and $\sum_{j \in [n_i]} \theta_{ij} = 1$, $j \in [n_i]$, $i \in [n]$. For a random sample $\boldsymbol{x}^{\mathrm{T}} = (\boldsymbol{x}_i^{\mathrm{T}})_{i \in [n]}$, where

$\boldsymbol{x}_i = (x_{ij})_{j\in[n_i]}^{\mathrm{T}}$ is the vector of the number of units that starts at stage w_i and go through the emanating edges, it holds that

$$f(\boldsymbol{x}\mid\boldsymbol{\theta}) \propto \prod_{i\in[n]}\prod_{j\in[n_i]} \theta_{ij}^{x_{ij}}, \tag{2.3.6}$$

where we further assumed that $\boldsymbol{x}_i \perp\!\!\!\perp \boldsymbol{x}_j \mid \boldsymbol{\theta}$, $i,j\in[n]$, and $\boldsymbol{\theta}^{\mathrm{T}} = (\boldsymbol{\theta}_i^{\mathrm{T}})_{i\in[n]}$. Therefore the likelihood is multinomial. Now assuming $\boldsymbol{\theta}$ has mutually independent parameters, the independence is retained a posteriori when the prior distribution is updated with the likelihood in equation (2.3.6). In particular this is conjugate if each stage probability vector $\boldsymbol{\theta}_i$ is given a Dirichlet prior distribution with parameter $\boldsymbol{a}_i = (a_{ij})_{j\in[n_i]}^{\mathrm{T}} \in \mathbb{R}_{>0}^{n_i}$. The posterior is then again Dirichlet with parameter $\boldsymbol{a}_i + \boldsymbol{x}_i$. Freeman and Smith [2011] proved that assuming mutual independence of the parameters and some other fairly mild conditions, the prior distribution is necessarily Dirichlet.

2.3.3 Dynamic Models

In this section we deal with models for processes that are observed at several points in time, usually called *time series*. The graphical models presented so far correspond to a fixed time, but now the random variables associated to the vertices of these will be allowed to vary in time, whilst, in most cases, the structure of the underlying graphical representation will remain constant through time. Such models allow for enough flexibility to represent situations where the relevant factors vary through time.

Let $\{\boldsymbol{Y}_t\}_{t\in[T]} = \{Y_i(t) : i \in [n]\}_{t\in[T]}$ be a n-dimensional time series with finite time horizon T, where $\{Y_i(t)\}_{t\in[T]}$, $i \in [n]$, is a univariate time series. We let $\mathcal{Y}_i$ and $\boldsymbol{\mathcal{Y}} = \bigtimes_{i\in[n]}\mathcal{Y}_i$ be the sample spaces of $Y_i(t)$ and $\boldsymbol{Y}(t)$ respectively, for any $t \in [T]$. The density function of a generic $Y_i(t)$ is parametrised by $\boldsymbol{\theta}_i(t)$ with sample space $\boldsymbol{\Theta}_i$. Let $\boldsymbol{Y}_A(t)^{\mathrm{T}} = (Y_i(t))_{i\in A}$, $\boldsymbol{Y}_A^t = (\boldsymbol{Y}_A(1)^{\mathrm{T}},\ldots,\boldsymbol{Y}_A(t)^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{Y}_{[n]}(t) = \boldsymbol{Y}(t)$ and $\boldsymbol{\theta}_A^t = (\boldsymbol{\theta}_A(1)^{\mathrm{T}},\ldots,\boldsymbol{\theta}_A(t)^{\mathrm{T}})^{\mathrm{T}}$, $A \subseteq [n]$. We denote with lower case letters instantiations of these random variables and vectors, and for any $t \in [T-1]$ the sample space of $(\boldsymbol{Y}(t),\boldsymbol{Y}(t+1))$ is $\boldsymbol{\mathcal{Y}}\times\boldsymbol{\mathcal{Y}}$. Finally, we denote with I^t the information set at time t, which includes all the relevant information available to the DM. We assume in this section that $I^t = \{I^{t-1},\boldsymbol{Y}(t-1)\}$, $t \in [T]_1$.

2.3.3.1 Dynamic Linear Models. Based on a distributed version of the Kalman Filter [Kalman, 1960], the theory of Dynamic Linear Models (DLMs) was introduced in Harrison and Stevens [1976] and described in the seminal book of West and Harrison [1997]. This theory allows for a coherent updating of probabilities in dynamic frameworks, which can be then embedded within the graphical models we introduce in the

$$\boldsymbol{Y}(t-1) \qquad \boldsymbol{Y}(t) \qquad \boldsymbol{Y}(t+1)$$

$$\longrightarrow \boldsymbol{\theta}(t-1) \longrightarrow \boldsymbol{\theta}(t) \longrightarrow \boldsymbol{\theta}(t+1) \longrightarrow$$
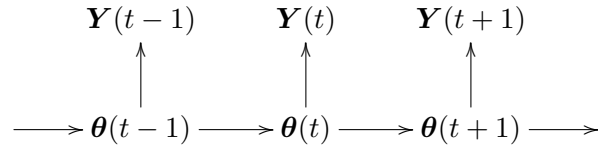
FIGURE 2.10: Conditional independence structure underlying the dynamic linear model class.

following sections. The key property of such models is the underlying conditional independence structure which assumes that at each time point all the relevant evidence is summarised in the distribution $\boldsymbol{\theta}(t) \mid \boldsymbol{\theta}(t-1)$, where $\boldsymbol{\theta}(t)^{\mathrm{T}} = (\boldsymbol{\theta}_i(t)^{\mathrm{T}})_{i \in [n]}$ has dimension r, and that consequently all the relevant information to predict $\boldsymbol{Y}(t)$ is synthesised in the distribution of $\boldsymbol{\theta}(t)$. This is depicted in Figure 2.10 implying, for each $t \in [T-1]_1$, $\boldsymbol{Y}(t) \perp\!\!\!\perp I^{t-1}, \boldsymbol{\theta}^{t-1} \mid \boldsymbol{\theta}(t)$, and $\boldsymbol{\theta}(t) \perp\!\!\!\perp I^{t-1} \mid \boldsymbol{\theta}(t-1)$.

We now introduce the general form of the Normal DLM.

Definition 2.3.35. *The* general Normal DLM *is defined by the following three equations.*

$$\boldsymbol{Y}(t) = F(t)^{\mathrm{T}}\boldsymbol{\theta}(t) + \boldsymbol{v}(t), \qquad\qquad \boldsymbol{v}(t) \sim \mathrm{N}\left(\boldsymbol{0}, V(t)\right), \qquad (2.3.7)$$

$$\boldsymbol{\theta}(t) = G(t)\boldsymbol{\theta}(t-1) + \boldsymbol{w}(t), \qquad\qquad \boldsymbol{w}(t) \sim \mathrm{N}\left(\boldsymbol{0}, W(t)\right), \qquad (2.3.8)$$

$$\boldsymbol{\theta}(1) \mid I^0 \sim \mathrm{N}\left(\boldsymbol{m}(0), C(0)\right),$$

where $V(t)$, $W(t)$ and $C(0)$ are known covariance matrices of dimensions $n \times n$, $r \times r$ and $r \times r$, respectively, $F(t)$ and $G(t)$ are generic matrices of dimensions $n \times r$ and $r \times r$, respectively, $\boldsymbol{m}(0) \in \mathbb{R}^n$ and the errors $\boldsymbol{v}(1), \dots, \boldsymbol{v}(T), \boldsymbol{w}(1), \dots, \boldsymbol{w}(T)$ are mutually independent, where $\boldsymbol{v}(t)$ and $\boldsymbol{w}(t)$, $t \in [T]$, have dimension n and r respectively,

Equation (2.3.7) is the *observation equation* specifying the distribution of $\boldsymbol{Y}(t)$ conditional on the *system vector* $\boldsymbol{\theta}(t)$, having mean $F(t)^{\mathrm{T}}\boldsymbol{\theta}(t)$ and variance $V(t)$, also called *observational variance matrix*. The matrix $F(t)$ is referred to as the *regression matrix*, whilst $\boldsymbol{v}(t)$ is the *observational error*. Equation (2.3.8) is the *system evolution equation*, which specifies how the values of the parameter vector evolve trough time. The matrices $G(t)$ and $W(t)$ are called respectively *system transfer* and *system variance* matrices. The vector $\boldsymbol{w}(t)$ is called *system error*.

Note that alternatively the general Normal DLM can be equivalently defined by substituting equations (2.3.7)-(2.3.8) with the following two conditional distributional specifications:

$$\boldsymbol{Y}(t) \mid \boldsymbol{\theta}(t) \sim \mathrm{N}\left(F(t)^{\mathrm{T}}\boldsymbol{\theta}(t), V(t)\right), \qquad \boldsymbol{\theta}(t) \mid \boldsymbol{\theta}(t-1) \sim \mathrm{N}\left(G(t)\boldsymbol{\theta}(t-1), W(t)\right).$$

Example 2.3.36. Special cases of the general Normal DLM model are:

- a *univariate DLM* is such that $\boldsymbol{Y}(t)$ has dimension 1 (it is therefore a scalar);

- a *linear univariate DLM* is such that $F(t) = (1, X_i(t))_{i \in [r-1]}^{\mathrm{T}}$, where $X_i(t)$ is a regression variable, $i \in [r-1]$;

- a univariate *regression* DLM is such that the entries of $F(t)$ are generic functions of $X_1(t), \ldots, X_{r-1}(t)$.

Although we do not focus in this thesis on modelling issues, we note here that the theory of DLMs comprises a large variety of methodologies to model for example both seasonal and polynomial temporal trends. The concept of *discount factors* is also widely used to elicit the values of the observational variance errors. Importantly, the conditional independence structure underlying DLMs provides a natural framework to *intervene* on the system by for example changing the value of some of the model's parameters. This strategy is usually implemented when the model has a poor forecasting performance.

The assumption of Gaussianity in Definition 2.3.35 is not strictly necessary but it entails some computational advantages. Under this assumption the updating equations of the parameters and the observables can be written in closed form and follow a Normal distribution, both in the univariate and the general case. The same property holds for the forecasting distributions, describing the behaviour of the system k-steps ahead in the future, for some $k \in \mathbb{Z}_{\geq 1}$. The associated recurrences when variances are known then reduce to the familiar Kalman Filter recurrences.

In practical applications the elicitation of the observational variance is critical and often prohibitive. However, there might be some information regarding its behaviour. Thus in practice $V(t)$ is often assumed unknown and a prior distribution is elicited. In the univariate case, an Inverse Gamma distribution can be given to this error, entailing a conjugate analysis as shown in Appendix C.2.1. It is then possible to obtain closed recurrences for both the updating and the forecasting distributions which, unconditionally on V(t), follow a T-distribution (see again Appendix C.2.1). Although we have shown in Appendix C.2.3 a conjugate analysis for the multivariate Normal model, this does not extend straightforwardly in a dynamic setting as in the univariate case to allow for sequential conjugate learning. We note here though that a variety of methods have been developed to approximate both numerically and analytically these recurrences [see West and Harrison, 1997]. Most importantly in the following section we introduce a class of multivariate DLMs that entertains exact closed form updating and forecasting routines.

2.3.3.2 Multiregression Dynamic Models. Queen and Smith [1993] introduced the class of *Multiregression Dynamic Models (MDMs)*, multivariate DLMs exhibiting a
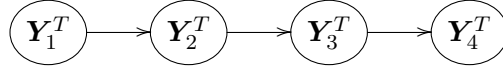
FIGURE 2.11: A directed acyclic graph associated to the conditional independence structure of a multiregression dynamic model.

conditional independence structure between the component time series which remains constant through time. Although each variable is modelled through a simple univariate regression DLM, where the regressors are specified by the conditional independence structure, the model class of MDMs is in general non Gaussian. The qualitative structure underlying an MDM can be represented through a DAG whose vertices are the component time series. Importantly, the well known BN model we reviewed in Section 2.3.2.2 is a special case of the MDM.

Definition 2.3.37. *An* MDM *for the time series* $\{\boldsymbol{Y}(t)\}_{t\in[T]}$ *is defined by a DAG* $\mathcal{G}$ *with vertex set* $\{Y_i^T : i \in [n]\}$ *together with the following n observation equations, system equation and initial information:*

$$Y_i(t) = \boldsymbol{F}_i(t)^{\mathrm{T}}\boldsymbol{\theta}_i(t) + v_i(t), \quad v_i(t) \sim (0, V_i(t)), \quad i \in [n];$$
$$\boldsymbol{\theta}(t) = G(t)\boldsymbol{\theta}(t-1) + \boldsymbol{w}(t) \quad \boldsymbol{w}(t) \sim (0, W(t));$$
$$\boldsymbol{\theta}(0) \mid I^0 \sim (\boldsymbol{m}(0), C(0)).$$

The system vector is $\boldsymbol{\theta}_i(t) \in \mathbb{R}^{r_i}$ *and* $\boldsymbol{F}_i(t)$, *of dimension* r_i, *is an arbitrary function of* $\boldsymbol{y}_{\Pi_i}^t$ *and* $\boldsymbol{y}_i^{t-1}$, *but not* $\boldsymbol{y}_{De_i}^t$ *and* $y_i(t)$. *The scalar observation variances* $V_i(t) \in \mathbb{R}_{>0}$, $i \in [n]$, *can be either known or unknown. The* $r \times r$ *matrices,* $r = \sum_{i=1}^n r_i$, $G(t) = \mathrm{blockdiag}(G_1(t), \ldots, G_n(t))$, $W(t) = \mathrm{blockdiag}(W_1(t), \ldots, W_n(t))$ *and* $C_0 = \mathrm{blockdiag}(C_1(0), \ldots, C_n(0))$, *are assumed known and such that* $W_i(t)$ *and* $C_i(t)$ *are covariance matrices, whilst* $G_i(t)$ *is a generic matrix, all of dimension* $s_i \times s_i$, $i \in [n]$.[5] *The errors* $v_1(t), \ldots, v_n(t), \boldsymbol{w}_1(t), \ldots, \boldsymbol{w}_n(t)$, *where* $\boldsymbol{w}_i(t) \sim (\boldsymbol{0}, W_i(t))$, *are mutually independent.*

Example 2.3.38. Consider a time series $\{\boldsymbol{Y}(t)\}_{t\in[T]} = \{Y_1(t), Y_2(t), Y_3(t), Y_4(t)\}_{t\in[T]}$. An MDM having the conditional independence structure depicted by the DAG in Figure 2.11 would have observation equations in which $\boldsymbol{F}_1(t)$ is a function of $\boldsymbol{y}_1^{t-1}$ and, for $i \in [4]_1$, $\boldsymbol{F}_i(t)$ is a function of y_{i-1}^t and y_i^{t-1}.

We now report from Queen and Smith [1993] two key results associated to MDMs. For ease of notation we assume the observational variances to be known, but these results straightforwardly generalise to the case of unknown observational variances.

Proposition 2.3.39. *For an MDM over a time series* $\{Y_i(t) : i \in [n]\}_{t\in[T]}$, *we have that* $\perp\!\!\!\perp_{i\in[n]}\boldsymbol{\theta}_i(t) \mid \boldsymbol{y}^{t-1}$ *and* $\boldsymbol{\theta}_i \perp\!\!\!\perp \boldsymbol{y}_{[n]\setminus Fa_i}^t \mid \boldsymbol{y}_{Fa_i}^t$.

---

[5]blockdiag denotes a block-diagonal matrix.

The first conditional independence in Proposition 2.3.39 indicates that the parameters associated to different component time series remain independent of each other through time. The second one guarantees that a parameter $\boldsymbol{\theta}_i(t)$, given the past observations of the variables with indices in the family set, is independent of the rest of the observed data. We show in Chapter 3 that these independences guarantee the MDM is an ideal model for the aggregation of expert judgements in the class of problems we address in this thesis.

Because of these results, we note that the overall updating of the multivariate time series can be performed locally for each of the component time series independently. Each of these follows, conditionally on the series with indices in the parent set, a simple univariate regression DLM. Therefore all the technology briefly reviewed in Section 2.3.3.1, as for example intervention and seasonal trends, can be directly transferred into MDMs.

Note that there is no assumption of Gaussianity in the definition of the MDM. However a *Normal MDM* can be defined such that the errors $v_t(i)$ and $\boldsymbol{w}_t$ follow a Gaussian distribution. In such cases the overall distribution is non Gaussian, but the forecasting and updating distributions can still be computed analytically and in closed form using the DLM machinery. Therefore MDMs, although being multivariate DLMs, do not need any approximated method. The following proposition formalises this fundamental observation.

Proposition 2.3.40. *For an MDM over a time series $\{Y_i(t) : i \in [n]\}_{t \in [T]}$, it holds that*

$$f(\boldsymbol{y}(t), \boldsymbol{\theta}(t) \mid I^{t-1}) = \prod_{i \in [n]} f(y_i(t) \mid \boldsymbol{y}_{\Pi_i}(t), \boldsymbol{\theta}_i(t), I^{t-1}) \pi(\boldsymbol{\theta}_i(t) \mid I^{t-1}). \qquad (2.3.9)$$

and

$$f\left(\boldsymbol{y}(t) \mid \boldsymbol{y}^{t-1}\right) = \prod_{i \in [n]} g_{t,i}\left(\boldsymbol{y}_{Fa_i}^t, \boldsymbol{\theta}_i(t)\right), \qquad (2.3.10)$$

where

$$g_{t,i} = \int_{\boldsymbol{\Theta}_i} f\left(y_i(t) \mid \boldsymbol{y}_{\Pi_i}^t, \boldsymbol{y}_i^{t-1}, \boldsymbol{\theta}_i(t)\right) \pi\left(\boldsymbol{\theta}_i(t) \mid \boldsymbol{y}_{Fa_i}^{t-1}\right) \mathrm{d}\boldsymbol{\theta}_t(i). \qquad (2.3.11)$$

Proposition 2.3.39 is a generalisation of the distributed learning property in BNs under the assumption of global independence we formalised in Proposition 2.3.13.

Example 2.3.41. For the MDM in Figure 2.11, equation (2.3.9) can be written as, letting $I = I^{t-1}$ for ease of notation,

$$f(\boldsymbol{y}(t), \boldsymbol{\theta}(t) \mid I) = f(y_1(t) \mid \boldsymbol{\theta}_1(t), I) \prod_{i \in [4]_1} f(y_i(t) \mid \boldsymbol{\theta}_i(t), y_{i-1}(t), I) \prod_{j \in [4]} \pi(\boldsymbol{\theta}_j(t) \mid I).$$

The terms in equation (2.3.11) appearing in the forecasting distribution of equation (2.3.10) can be written for this example as

$$
g_{t,i} = \begin{cases} \int_{\boldsymbol{\Theta}_i} f\left(y_i(t) \mid \boldsymbol{y}_i^{t-1}, \boldsymbol{\theta}_i(t)\right) \pi\left(\boldsymbol{\theta}_i(t) \mid \boldsymbol{y}_i^{t-1}\right) \mathrm{d}\boldsymbol{\theta}_i(t), & i = 1, \\ \int_{\boldsymbol{\Theta}_i} f\left(y_i(t) \mid \boldsymbol{y}_i^{t-1}, \boldsymbol{y}_{i-1}^{t}, \boldsymbol{\theta}_i(t)\right) \pi\left(\boldsymbol{\theta}_i(t) \mid \boldsymbol{y}_i^{t-1}, \boldsymbol{y}_{i-1}^{t-1}\right) \mathrm{d}\boldsymbol{\theta}_i(t), & i \in [4]_1. \end{cases}
$$

We now introduce a special case of the MDM class, the *linear MDM* [Queen and Smith, 1993], which is very simple to work with.

Definition 2.3.42. *A* Linear Multiregression Dynamic Model (LMDM) *is an MDM where the errors are assumed to be Gaussian and each component time series is modelled as a univariate linear DLM.*

In an LMDM, at each time point, each $Y_i(t)$ is modelled as a linear regression with the contemporaneous parent variables as regressors. Therefore the LMDM is a dynamic extension of the Gaussian BN model introduced in Proposition 2.3.15.

Example 2.3.43. The observation equations of an LMDM respecting the DAG in Figure 2.11 (and using an obvious generalisation of the notation in equation (2.3.4)) are as follows

$$
Y_1(t) = \theta_{01}(t) + v_1(t), \qquad Y_2(t) = \theta_{02}(t) + \theta_{12}(t)Y_1(t) + v_2(t),
$$
$$
Y_3(t) = \theta_{03}(t) + \theta_{23}(t)Y_2(t) + v_3(t), \qquad Y_4(t) = \theta_{04}(t) + \theta_{34}(t)Y_3(t) + v_4(t).
$$

Queen and Smith [1993] and Queen et al. [2008] described methods to analytically compute the first two moments of $\boldsymbol{Y}(t)$ under the assumption of an LMDM. These simply consist of a sequential use of the tower properties of the first two moments. Many of the results we present in the following chapters use similar techniques.

Because of its independence properties and its flexibility, the MDM has now been successfully applied in practice in many diverse domains: traffic flows [Anacleto et al., 2013], biology [Oates, 2013], brain connectivity [Costa et al., 2015], brand sales [Queen, 1994] and finance [Zhao and West, 2015]. Furthermore, a wide tool-kit of more advanced modelling techniques has now been implemented within MDMs to allow for causal reasoning [Queen and Albers, 2009], model choice and averaging [Costa et al., 2015, Zhao and West, 2015] and heteroscedasticity [Anacleto et al., 2013].

2.3.3.3 Dynamic Chain Graphs. Although the MDM allows for great flexibility in modelling multivariate dynamic domains, its underlying DAG structure implies the presence of directional associations only and does not allow for any symmetric relationship.
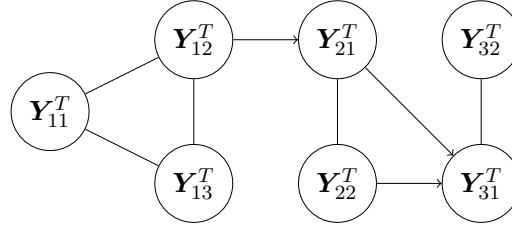
FIGURE 2.12: Dynamic variant of the chain graph model in Figure 2.5.

To address this issues, Queen and Smith [1992] and Anacleto and Queen [2013] developed the *Dynamic Chain Graph (DCG)* model, a multivariate DLM whose underlying conditional independence structure can be represented by a CG.

For the purpose of this section only, let $\boldsymbol{Y}(t)$ be partitioned into N vector time series of dimensions $n_1, \ldots, n_N$, $\sum_{i \in [N]} n_i = n$, so that $\boldsymbol{Y}(t)^{\mathrm{T}} = (\boldsymbol{Y}_i(t)^{\mathrm{T}})_{i \in [N]}$ where $\boldsymbol{Y}_i(t)^{\mathrm{T}} = (Y_{ij}(t))_{j \in [n_i]}$. Let $\boldsymbol{Y}_{ij}^t = (Y_{ij}(1), \ldots, Y_{ij}(t))^{\mathrm{T}}$. Suppose the independence structure is such that any two $\boldsymbol{Y}_{ij}^T$ and $\boldsymbol{Y}_{ik}^T$ are connected by an undirected edge in the underlying CG, $j, k \in [n_i]$, $i \in [N]$.

Definition 2.3.44. *A* DCG *consists of a CG* $\mathcal{G}$, $V(\mathcal{G}) = \{Y_{ij}^T : i \in [N], j \in [n_i]\}$, *together with the following N observations equations, system equation and initial information:*

$$\boldsymbol{Y}_i(t) = \boldsymbol{F}_i(t)^{\mathrm{T}}\boldsymbol{\theta}_i(t) + v_i(t), \quad \boldsymbol{v}_i(t) \sim (\boldsymbol{0}, V_i(t)), \quad i \in [N],$$
$$\boldsymbol{\theta}(t) = G(t)\boldsymbol{\theta}(t-1) + \boldsymbol{w}(t), \quad \boldsymbol{w}(t) \sim (\boldsymbol{0}, W(t)),$$
$$\boldsymbol{\theta}(1) \mid I(0) \sim (\boldsymbol{m}(0), C(0)).$$

The vector $\boldsymbol{F}_i(t)^{\mathrm{T}} = \left(\boldsymbol{F}_{ij}(t)^{\mathrm{T}}\right)_{j \in [n_i]}$ *includes the subvectors* $\boldsymbol{F}_{ij}(t) \in \mathbb{R}^{r_{ij}}$ *known functions of* $\boldsymbol{y}_i^{t-1}$ *and* $\boldsymbol{y}_{\Pi_{ij}}^t$, *where* Π_{ij} *is the set of the indices of the parents of* $\boldsymbol{Y}_{ij}^T$ *in* $\mathcal{G}$, $j \in [n_i]$. *The system vector* $\boldsymbol{\theta}(t)^{\mathrm{T}} = (\boldsymbol{\theta}_i(t)^{\mathrm{T}})_{i \in [N]} \in \mathbb{R}^r$, *where* $\boldsymbol{\theta}_i(t) \in \mathbb{R}^{r_i}$ *is the system vector of* $\boldsymbol{Y}_i(t)$, *with* $r_i = \sum_{j \in [n_i]} r_{ij}$ *and* $r = \sum_{i \in [N]} r_i$. *The $n_i \times n_i$ covariance matrix* $V_i(t)$ *is the observational variance for* $\boldsymbol{Y}_i(t)$. *The $r \times r$ matrices* $G(t) = \mathrm{blockdiag}(G_1(t), \ldots, G_N(t))$, $W(t) = \mathrm{blockdiag}(W_1(t), \ldots, W_N(t))$ *and* $C(0) = \mathrm{blockdiag}(C_1(0), \ldots, C_N(0))$ *are assumed known and such that* $W_i(t)$ *and* $C_i(0)$ *are* $r_i \times r_i$ *covariance matrices and* $G_i(t)$ *is a generic matrix. The errors* $\boldsymbol{v}_1(t), \ldots, \boldsymbol{v}_N(t), \boldsymbol{w}_1(t), \ldots, \boldsymbol{w}_N(t)$, $\boldsymbol{w}_i(t) \sim \mathrm{N}(\boldsymbol{0}, W_i(t))$, *are assumed mutually independent.*

Example 2.3.45. Consider the DCG defined by the CG of Figure 2.12. Since the topology of this CG is the same as the one in Figure 2.5, this DCG has three strong components, corresponding, for each time slice $t \in [T]$, to $\boldsymbol{Y}_1(t) = (Y_{11}(t), Y_{12}(t), Y_{13}(t))^{\mathrm{T}}$, $\boldsymbol{Y}_2(t) = (Y_{21}(t), Y_{22}(t))^{\mathrm{T}}$ and $\boldsymbol{Y}_3(t) = (Y_{31}(t), Y_{32}(t))^{\mathrm{T}}$. Note that in this case we have $\boldsymbol{Y}(t)^{\mathrm{T}} = (\boldsymbol{Y}_1(t)^{\mathrm{T}}, \boldsymbol{Y}_2(t)^{\mathrm{T}}, \boldsymbol{Y}_3(t)^{\mathrm{T}})$, as confirmed by Figure 2.6. The observation equations are specified by the following vectors: $\boldsymbol{F}_1(t)^{\mathrm{T}} = (\boldsymbol{F}_{11}(t)^{\mathrm{T}}, \boldsymbol{F}_{12}(t)^{\mathrm{T}}, \boldsymbol{F}_{13}(t)^{\mathrm{T}})$, functions of
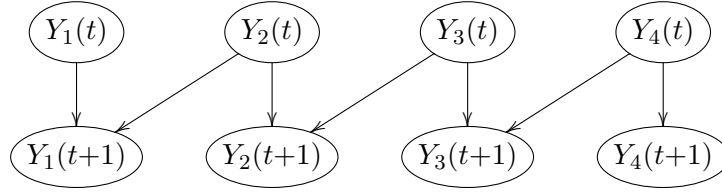
FIGURE 2.13: Example of a 2-time-slice Bayesian network for a multivariate time series comprising four univariate series.

$\boldsymbol{y}_1^{t-1}$; $\boldsymbol{F}_2(t)^{\mathrm{T}} = (\boldsymbol{F}_{21}(t)^{\mathrm{T}}, \boldsymbol{F}_{22}(t)^{\mathrm{T}})$, functions of $\boldsymbol{y}_2^{t-1}$, where $\boldsymbol{F}_{21}(t)$ is also a function of $\boldsymbol{y}_{13}^t$; $\boldsymbol{F}_3(t)^{\mathrm{T}} = (\boldsymbol{F}_{31}(t)^{\mathrm{T}}, \boldsymbol{F}_{32}(t)^{\mathrm{T}})$, function of $\boldsymbol{y}_3^{t-1}$, where $\boldsymbol{F}_{31}(t)$ is also a function of $\boldsymbol{y}_2^t$.

It can be shown that the results in Propositions 2.3.39 and 2.3.40 for MDMs hold for DCGs if applied to the chain component observation series $\boldsymbol{Y}_i(t)$ and their associated system vectors $\boldsymbol{\theta}_i(t)$, $i \in [N]$. Therefore in DCGs the sequential updating and forecasting of the relevant probabilities can be distributed across the chain components of the underlying CG.

2.3.3.4 Dynamic Bayesian Networks. The two described multivariate models are not as commonly used in practice as the *Dynamic Bayesian Network (DBN)* model class [Murphy, 2002]. DBNs extend the BN framework to dynamic and stochastic domains. We discuss these models in this thesis to provide an example of a model class whose conditional independence structure is not retained as time progresses and therefore do not enjoy closed form updating routines.

For the purpose of this thesis, and as often in practice, we consider only stationary, feed-forward DBNs respecting the first order Markov assumption [see e.g. Oates, 2013]. These DBNs can be simply described by an initial distribution over the first time point and a BN having as vertex set two generic time slices. Such latter BN is usually called 2-Time slice Bayesian Network (2-TBN).

Definition 2.3.46. *A* 2-TBN *for* $\{\boldsymbol{Y}(t)\}_{t \in [T]}$ *is a BN with DAG* $\mathcal{G}$ *such that, fixed a* $t \in [T-1]$, $V(\mathcal{G}) = \{Y_i(t), Y_i(t+1) : i \in [n]\}$, *any vertex* $Y_i(t)$ *has no parents and there are no edges* $(Y_i(r), Y_j(r))$, $i, j \in [n]$, $r = t, t+1$.

Definition 2.3.47. *A* DBN *for the time series* $\{\boldsymbol{Y}(t)\}_{t \in [T]}$ *is a pair* $(\mathcal{G}, \mathcal{G}')$, *such that* $\mathcal{G}$ *is a BN with vertex set* $V(\mathcal{G}) = \{Y_i(1) : i \in [n]\}$, *and* $\mathcal{G}'$ *is a 2-TBN such that its vertex set* $V(\mathcal{G}')$ *is equal to* $\{Y_i(t), Y_i(t+1) : i \in [n]\}$.

It is therefore straightforward to notice that a recursive formula for the density function of a DBN similar to the one for BNs in Lemma 2.3.6 exists. This is because a DBN can be thought of as the concatenation of BNs.
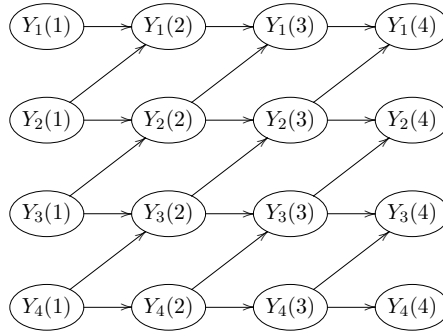
FIGURE 2.14: Unrolled version of the dynamic Bayesian network of Example 2.3.48 with 2-time-slice Bayesian network of Figure 2.13.

Example 2.3.48. The BN in Figure 2.13 is a valid 2-TBN because its topology respects the conditions of Definition 2.3.46. Suppose a DBN has such 2-TBN and the BN associated to the initial distribution over $\boldsymbol{Y}(1)$ has empty edge set, i.e. represents the independence model in Definition 2.3.4. Then its probability density function factorises as (suppressing the dependence on the parameter vector for ease of notation)

$$f\left(\boldsymbol{y}^T\right) = \prod_{i\in[4]} f(y_i(1)) \prod_{t\in[T]_1} f(y_4(t) \mid y_4(t-1)) \prod_{j\in[3]} f(y_j(t) \mid y_j(t-1), y_{j+1}(t-1))$$

Although DBNs entail an effective recursive factorisation of the associated density function, thus requiring a low number of probabilities to be elicited in order to fully specify the model, inference in such models cannot be performed as easily as in both BNs and MDMs. This is because the initial conditional independence structure is broken through time. As shown in Figure 2.14 [from Boyen and Koller, 1998], the so-called unrolled version of the DBN of Example 2.3.48, after a small number of time steps all the variables in a same time slice become correlated with each other. Independence can only be retained if each component time series is independent of the others, i.e. if the 2-TBN has edges of the type $(Y_i(t), Y_i(t+1))$ and an initial distribution described by an independence model. Therefore the efficient and distributed recursions for both inference and forecasting associated to MDMs and DCGs do not transfer to generic DBNs. This is the reason why DBNs are not an effective model class for exact aggregation methods in the problems we address in Chapter 3. However, we note here that a variety of methodologies based on stochastic approximated methods have been developed for tractable inferences in such models [Boyen and Koller, 1998, Koller and Lerner, 2001].

2.3.4 Causality

Statistical reasoning has historically largely avoided investigating causal inferences. Classical methods, as for example regression analysis, are in general only able to predict new

values of a variable given the values of a set of covariates, and how these relate to the response variable. Until recently, there has not been a formal study of causal relationships between sets of random variables to assess, for example, whether a variable can be considered as a *cause* for another one or not. However, artificial intelligence researchers such as Pearl [2000] and Spirtes et al. [1993], provided a systematic study of the causal mechanisms underlying a given hypothesised BN model when certain variables are *manipulated* and fixed to take a certain value in their sample space. This theory supposes the existence of an underlying idle system, i.e. one where its variables are observed and not manipulated to a certain value.

We introduce here the concept of causality since, for the purpose of decision support, it is often helpful to match data coming from experiments in a laboratory, where variables are held fixed to a particular value, to observational one, to provide a revised and more focused specification of the relevant probabilities.

Shafer [1996] discusses causal reasoning for models depicted by trees. As also suggested by Smith [2010], trees are particularly suitable to represent causal relationships, since these can uniquely describe the actual narrative underlying the unfolding of events. We also note here that causal reasoning is more tenable for Bayesian decision analysis than in pure inferential reasoning since DMs need to think hard on how the problem at hand unfolds when eliciting both their probabilistic and preferential beliefs. Therefore, a DM's model can often be used to answer various causal questions [Smith, 2010].

We now formally define the concept of intervention as introduced in Pearl [2000].

Definition 2.3.49. *A* Perlean intervention *on* $\boldsymbol{Y}_A$, $A \subset [n]$, *consists of fixing these variables to a (known) value* $\boldsymbol{y}_A$. *The resulting density is written* $f(\boldsymbol{y} \mid (\boldsymbol{Y}_A = \boldsymbol{y}_A))$.

As noted in Section 2.3.2.2, there are several BNs entertaining the same density factorisation and therefore leading to the same kind of inferences. Causal reasoning cannot thus be straightforwardly followed in generic BNs. We define here a particular class of BNs that are causal in the Perlean sense.

Definition 2.3.50. *A BN* $\mathcal{G}$ *is a* Causal Bayesian Network (CBN *if, for any Perlean intervention on* $\boldsymbol{Y}_A$, $A \subset [n]$, *it holds that*

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}, (\boldsymbol{Y}_A = \boldsymbol{y}_A)) = \prod_{i \in [n] \setminus A} f(y_i \mid \boldsymbol{\theta}_i, \boldsymbol{y}_{\Pi_i}).$$

Example 2.3.51. Consider the CBN with DAG in Figure 2.1. Suppose we intervene on this DAG and set $Y_2 = y_2$. The resulting factorisation is equal to

$$f(y_1, y_2, y_3, y_4 \mid \boldsymbol{\theta}, do(Y_2 = y_2)) = f(y_1 \mid \boldsymbol{\theta}_1) f(y_3 \mid y_1, y_2, \boldsymbol{\theta}_3) f(y_4 \mid y_1, \boldsymbol{\theta}_4).$$
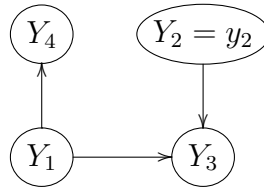
FIGURE 2.15: Causal version of the Bayesian Network in Figure 2.1 under the intervention $Y_2 = y_2$.

Such intervention can be depicted graphically by deleting every edge into Y_2, in this case simply (Y_1, Y_2), and changing the label of the vertex associated to the intervened variable to $Y_2 = y_2$. Such DAG is reported in Figure 2.15.

Definition 2.3.52. *For an intervention* $(\boldsymbol{Y}_A = \boldsymbol{y}_A)$ *over the variables of a CBN* $\mathcal{G}$, $V(\mathcal{G}) = \{Y_i : i \in [n]\}$, *we call* manipulated DAG $\mathcal{G}'$ *the network with edge set* $\mathbb{E}(\mathcal{G}') = \mathbb{E}(\mathcal{G}) \setminus \{(Y_i, Y_j) \in E(\mathcal{G}) : j \in A\}$ *and* $V(\mathcal{G}') = \{Y_i : i \in \{[n] \setminus A\}\} \cup \{Y_i = y_i : i \in A\}$.

If a BN is believed to be causal, then experimental evidence can be used to update the parameter densities as formalised below.

Proposition 2.3.53. *Suppose global independence holds and assume the experimental sample* $\boldsymbol{x}$ *from the same population as* $\boldsymbol{Y}$ *is ancestral with respect to the manipulated DAG for the intervention* $do(\boldsymbol{Y}_A = \boldsymbol{y}_A)$. *Let I be the set of indices of both the vertices whose children are not sampled and the leaves of the DAG. Then, for a CBN the posterior factorises as*

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) = \prod_{i \in I} \prod_{\substack{j \in A_i \\ j \notin A}} \pi(\boldsymbol{\theta}_j \mid \boldsymbol{x}) \prod_{k \notin \{A_i \cup A\}} \pi(\boldsymbol{\theta}_k).$$

The concept of causality for BNs has been extended in Daneshkhah and Smith [2004] to take into account manipulations of the parameter vector, usually called *randomised interventitions* [see e.g. Daneshkhah and Smith, 2004, Lauritzen, 2001]. These consist of fixing a random parameter of the density of a random variable to take a particular known value. Daneshkhah and Smith [2004] defined the concept of an hypercausal BN, which is one where certain randomised interventions respect the topology of the DAG. Most importantly they showed that local and global independence holds iff the BN is hypercausal and therefore also in the case the model is a CBN. Thus the validity of the causal assumption can be checked through reasoning about the faithfulness of global and local independence and vice versa. In Chapter 3 below we define a new class of randomised interventions related to committing to a countermeasure policy. We are able to show that Perlean interventions in CBNs can be seen as a special case of our methodology.

We have focused here on causality over graphical directed structures, since these naturally provide a framework for causal reasoning. In these models variables are ordered

and there are no symmetric relationships. This is the reason why causal arguments in MN models are much more difficult to develop [see e.g. Lauritzen and Richardson, 2002]. Although much of the literature on statistical causality centred on the BN model class, the causal semantic has been extended to a variety of frameworks, e.g. CEGs [Thwaites, 2013, Thwaites et al., 2010], CGs [Lauritzen and Richardson, 2002], DBNs [Eichler and Didelez, 2007], IDs [Dawid, 2002], MDMs [Queen and Albers, 2009] and others [Aalen et al., 2012, Dawid and Didelez, 2010, Smith and Figueroa, 2007]. We point out that Dawid and Didelez [2010] developed causal arguments in the framework of *dynamic treatment strategies* [Murphy, 2003]. These are dynamic models in medical contexts where the objective is to identify the causal effect of a particular treatment. The computation of these effects is based on backward inductive arguments, which mirror the recursions we develop in Chapter 4 for IDSSs.

2.3.5   Emulators

Although probabilistic reasoning is now widespread in many areas of science, deterministic modelling is still often performed in practice as noted in Chapter 1. Such deterministic models usually consist of huge simulators using approximate methods to numerically solve big systems of differential equations describing some natural process. This is often the case in climate change modelling for example [Rougier and Crucifix, 2014].

Even if we believed these simulator models were true - which would be heroic - for coherent Bayesian analyses to use these, there is the need to define a probability distribution over the corresponding space of possibilities. This is because unmodelled uncertainty appears at various stages of the deterministic computations of simulators, as extensively discussed in Kennedy and O'Hagan [2000]. Probability distributions are then usually achieved by building an **emulator** over their outputs. The literature on emulators is now very vast [Kennedy and O'Hagan, 2001, Kennedy et al., 2006, O'Hagan, 2006, Santner et al., 2003] and for the purpose of this thesis we focus here only on Bayesian methods. Within the Bayesian literature a variety of methodologies have been developed to account for both large-dimensional and dynamic outputs [Conti et al., 2009, Craig et al., 2001, Goldstein and Rougier, 2006, Liu and West, 2009, Rougier, 2008, Williamson and Goldstein, 2011].

2.3.5.1   Modelling of Computer Outputs. A simulator is a function $g(\cdot)$ that maps inputs $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$, for some arbitrary space $\boldsymbol{\mathcal{Z}}$, into an output $y = g(\boldsymbol{z})$. In its vanilla form the output y is univariate and constant through time. Often in practice only a small set of training runs at inputs $\boldsymbol{z}_1, \dots, \boldsymbol{z}_N$ of the simulator are available, whose

outputs $y_1 = g(\boldsymbol{z}_1), \ldots, y_N = g(\boldsymbol{z}_N)$ are observed and treated as data. Only a small number of such outputs can be observed since simulators are usually slow and each single evaluation can take weeks if not months. An emulator is then an approximation $\hat{g}(\cdot)$ of $g(\cdot)$ such than at an input point $\boldsymbol{z}_i$, $\hat{g}(\boldsymbol{z}_i) = g(\boldsymbol{z}_i)$, whilst at other points it consists of a distribution whose mean represents a plausible interpolation of the training runs and its variance the uncertainty associated to such interpolation.

2.3.5.2 Gaussian Process Modelling. The most common modelling technique is to assume that $g(\cdot)$ behaves as a *Gaussian Process (GP)*. Formally, $g(\cdot)$ has a GP distribution if, for every $N \in \mathbb{Z}_{\geq 1}$, the joint distribution of $g(\boldsymbol{z}_1), \ldots, g(\boldsymbol{z}_N)$ is multivariate Normal for all $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N \in \boldsymbol{\mathcal{Z}}$. The distribution of the GP is therefore characterised by its mean $m(\boldsymbol{z}) = \mathbb{E}(g(\boldsymbol{z}))$ and its covariance function $c(\boldsymbol{z}, \boldsymbol{z}') = \mathbb{C}(g(\boldsymbol{z}), g(\boldsymbol{z}'))$, for $\boldsymbol{z}, \boldsymbol{z}' \in \boldsymbol{\mathcal{Z}}$. In general, $m(\cdot)$ may be any function, but $c(\cdot, \cdot)$ is non negative definite for every $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N \in \boldsymbol{\mathcal{Z}}$ and any $N \in \mathbb{Z}_{\geq 1}$. The mean and covariance functions are usually modelled hierarchically as $g(\boldsymbol{z}) = m(\boldsymbol{z}) + e(\boldsymbol{z}) = \boldsymbol{h}(\boldsymbol{z})^{\mathrm{T}} \boldsymbol{\beta} + e(\boldsymbol{z})$, where $\boldsymbol{h}(\boldsymbol{z}) = (h_i(\boldsymbol{z}))_{i \in [p]}^{\mathrm{T}}$ is a vector of p known functions, $\boldsymbol{\beta} = (\beta_i)_{i \in [p]}^{\mathrm{T}}$ is a vector of p unknown coefficients and $e(\boldsymbol{z})$ is a mean-zero GP with covariance $c(\cdot, \cdot)$. The vector $\boldsymbol{h}$ often simply consists of simple monomial functions of $\boldsymbol{z}$. The covariance is usually defined as $c(\boldsymbol{z}, \boldsymbol{z}') = \psi r(\boldsymbol{z} - \boldsymbol{z}')$, where $r(\cdot)$ is a correlation function such that $r(0) = 1$, $\boldsymbol{z}, \boldsymbol{z}' \in \boldsymbol{\mathcal{Z}}$, and $\psi \in \mathbb{R}_{>0}$. This choice implies a stationary process since the correlation only depends on the distance between two points. A variety of correlation functions have been defined in the literature and these are often used in geostatistical modelling [see e.g. Diggle and Ribeiro, 2007]. The model definition is then completed by eliciting a prior distribution over the parameters $\boldsymbol{\beta}$ and ψ. Often a weak improper prior is given to such parameters, but conjugate analysis can be performed by choosing appropriate prior distributions just as in the Normal linear model case [see e.g. Haylock and O'Hagan, 1996].

Once an emulator is built, using for example the GP structure we exemplified here, each evaluation of the simulator can be used to update the probability distribution of the emulator in a Bayesian fashion. Furthermore, additional uncertainty measures can be included in the modelling. For example, calling y_i' the true value of the system the simulator is modelling, given observed inputs $\boldsymbol{z}_i$, one could set $y_i' = g(\boldsymbol{z}_i) + v$, where v is some error following for example a Gaussian distribution [see Kennedy and O'Hagan, 2000, for more comments on these modelling techniques].

2.4 Utility Models

The other main ingredient of the SEU model is the utility function $u(\boldsymbol{r}, d)$. In this section we provide a broad overview of utility theory with a particular focus to the multiattribute case, i.e. when $\boldsymbol{R}$ is large dimensional. Let $\boldsymbol{R} = (R_i)_{i \in [m]}^{\mathrm{T}}$ and $\boldsymbol{r}$ and r_i be instantiations of $\boldsymbol{R}$ and R_i taking values in $\boldsymbol{\mathcal{R}}$ and $\mathcal{R}_i$ respectively. Note that in the notation of Section 2.2, each R_i can correspond to random quantities, decisions, or a combination of the two.

Here we first introduce independence concepts for preferences mirroring those for probabilistic reasoning. We then show how particular sets of independence assumptions can lead to classes of utility factorisations that, just as in the probabilistic case, can highly reduce the burden of utility elicitation. We then consider fairly recent utility factorisations that arise from underlying graphical models depicting preferential indifferences. We conclude the section with a review of utility theory for single attributes which, under the assumption of a particular utility factorisation, comes often into play in multivariate settings.

Before that, we briefly summarise important aspects of utility theory that we do not deal with in this thesis since these are not central for our results. Firstly, we do not discuss value functions, although these are widely used in practice [Belton and Stewart, 2002, Keeney and Raiffa, 1976]. There are two main reasons for this: one, the coherence and the rationale of the Bayesian formalism is retained only using utility functions; two, value functions, differently from utilities, are not able to represent attitudes towards risk.

Secondly, we assume in this thesis that the vector $\boldsymbol{R}$ includes a *complete* and *non-redundant* collection of attributes so that this covers all the important aspects of the problem and each factor is not double counted. In addition each R_i has to be *operational* so that it can be meaningfully used in the analysis and the DM can provide true preferential assessments about it. More broadly, we assume that the vector $\boldsymbol{R}$ includes all the relevant factors of the system under study and the DM can uniquely provide preferential assessments over its elements. In practice to identify such a vector of attributes an *objective tree* [Keeney and Raiffa, 1976] is built, which breaks down each attribute into sub-attributes until the leaves of the tree consist of operational attributes only. An example of the objective tree elicited during the Chernobyl project is presented in Figure 2.16. The root of this tree is the main factor related to a nuclear emergency: how the accident affects humans' living condition. This is then split in two sub-attributes: one concerning the effects of the accident, the second the use of resources. Effects is again split in two sub-attributes and the process continues in this fashion until a vertex
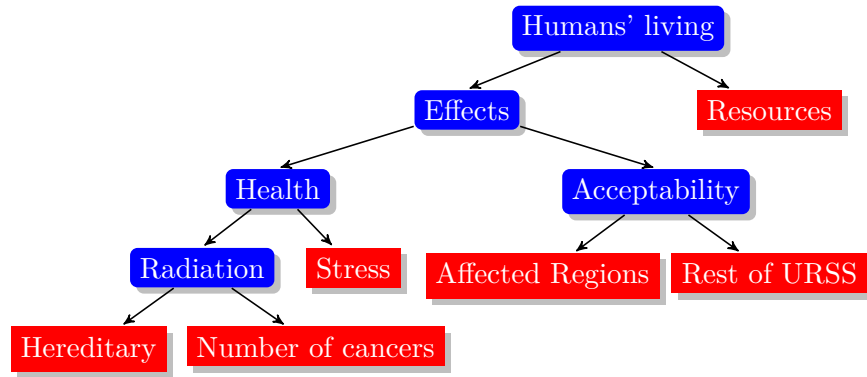
FIGURE 2.16: Objective tree elicited at the end of the Chernobyl project from Papamichail and French [2013]. The vertices in red of the tree are leaves, whilst the blue ones are positions.

represents an operational attribute. Each leaf of this tree would then correspond in our notation to an element R_i of $\boldsymbol{R}$.

Finally, we do not discuss techniques to assess and elicit utility functions, either their algebraic form or their features [see e.g. Clemen, 1996, Keeney, 1992, Keeney and Raiffa, 1976, Von Winterfeldt and Edwards, 1986]. These are however fundamental in any formal decision analysis to faithfully represent the preferences of DMs.

2.4.1 Independence Concepts

As the dimension of $\boldsymbol{R}$ might be arbitrarily large, a variety of independence concepts have been introduced in the literature to both simplify the utility elicitation and describe various sets of indifferences. We introduce a few of these following Keeney and Raiffa [1976].

Definition 2.4.1. *An attribute $\boldsymbol{R}_A$ is said to be **Utility Independent (UI)** of $\boldsymbol{R}_B$, for a partition A, B of $[m]$,[6] if the utility for $\boldsymbol{R}_A$ does not change when the attributes in $\boldsymbol{R}_B$ are varied.*

Proposition 2.4.2. *Under the conditions of Definition 2.4.1, if $\boldsymbol{R}_A$ is UI of $\boldsymbol{R}_B$ then $u(\boldsymbol{r}) = a(\boldsymbol{r}_B) + b(\boldsymbol{r}_B)u(\boldsymbol{r}_A)$, where $a(\cdot)$ and $b(\cdot) > 0$ depend on $\boldsymbol{r}_B$ and not on $\boldsymbol{r}_A$.*

In order to understand the meaning of this independence, consider the following example. If the DM believes that the prevalence of tumours is UI of the amount of contamination, then the utility function describing the prevalence of tumours does not depend on the level of contamination of the area. For example, the utility of having a low number of cancer cases would be the same both when the area is highly contaminated and when there is no radiation at all. Note that utility independence is not necessarily symmetric:

---

[6]More generally, a partition $B_1, \ldots, B_n$ of a set $[m]$ is such that $\cup_{i \in [n]} B_i = [m]$ and $\cap_{i \in [n]} B_i = \emptyset$.

if the prevalence of tumours is UI of the level of contamination then it does not follow that the converse is true.

We now consider two particular sets of utility independence statements.

Definition 2.4.3. *We say that $\boldsymbol{R}$ has* singly UI attributes *if R_i UI $\boldsymbol{R}_{[m]\setminus\{i\}}$, for every $i \in [m]$.*

Definition 2.4.4. *We say that $\boldsymbol{R}$ has* mutually UI attributes *if, for every $A \subseteq [m]$, $\boldsymbol{R}_A$ UI $\boldsymbol{R}_{[m]\setminus A}$.*

We now consider a generalisation of utility independence [see e.g. Abbas, 2010].

Definition 2.4.5. *An attribute $\boldsymbol{R}_A$ is said to be* **Conditional Utility Independent (CUI)** *of $\boldsymbol{R}_B$ given $\boldsymbol{R}_C$, for a partition A, B, C of $[m]$, if the utility of $\boldsymbol{R}_A$ does not change when the attributes in $\boldsymbol{R}_B$ are varied, for each instantiation of $\boldsymbol{R}_C$.*

Proposition 2.4.6. *Under the conditions of Definition 2.4.5, if $\boldsymbol{R}_A$ is CUI of $\boldsymbol{R}_B$ given $\boldsymbol{R}_C$ then*

$$u(\boldsymbol{r}) = a(\boldsymbol{r}_B, \boldsymbol{r}_C) + b(\boldsymbol{r}_B, \boldsymbol{r}_C)h(\boldsymbol{r}_A, \boldsymbol{r}_C) \tag{2.4.1}$$

for some a, h and $b > 0$, functions of their arguments only.

In order to give an explicit form to the function h of equation (2.4.1) in terms of utilities, we introduce a generalisation of the utility function of Definition 2.2.1, which more easily can represent CUI statements. Let $\boldsymbol{r}_A^* = (r_i^*)_{i\in A}^{\mathrm{T}}$ and $\boldsymbol{r}_A^0 = (r_i^0)_{i\in A}^{\mathrm{T}}$, $A \subseteq [n]$, where r_i^* and r_i^0 are respectively the best and the worst outcome of attribute R_i.

Definition 2.4.7. *The* normalised conditional utility *function for $\boldsymbol{R}_A$ given $\boldsymbol{R}_{[m]\setminus A}$ is defined as*

$$u(\boldsymbol{r}_A \mid \boldsymbol{r}_{[m]\setminus A}) = \frac{u(\boldsymbol{r}_A, \boldsymbol{r}_{[m]\setminus A}) - u(\boldsymbol{r}_A^0, \boldsymbol{r}_{[m]\setminus A})}{u(\boldsymbol{r}_A^*, \boldsymbol{r}_{[m]\setminus A}) - u(\boldsymbol{r}_A^0, \boldsymbol{r}_{[m]\setminus A})}.$$

We also define the normalised conditional disutility *function as*

$$\check{u}(\boldsymbol{r}_A \mid \boldsymbol{r}_{[m]\setminus A}) = 1 - u(\boldsymbol{r}_A \mid \boldsymbol{r}_{[m]\setminus A})$$

From e.g. Abbas [2010] we have the following.

Proposition 2.4.8. *Under the conditions of Definition 2.4.5, if $\boldsymbol{R}_A$ CUI $\boldsymbol{R}_B$ given $\boldsymbol{R}_C$ then*

$$u(\boldsymbol{r}_A \mid \boldsymbol{r}_B, \boldsymbol{r}_C) = u(\boldsymbol{r}_A \mid \boldsymbol{r}_B^0, \boldsymbol{r}_C) = u(\boldsymbol{r}_A \mid \boldsymbol{r}_B^*, \boldsymbol{r}_C).$$

Furthermore, the terms in equation (2.4.1) can be written as $a(\boldsymbol{r}_B, \boldsymbol{r}_C) = u(\boldsymbol{r}_A^0, \boldsymbol{r}_B, \boldsymbol{r}_C)$, $b(\boldsymbol{r}_B, \boldsymbol{r}_C) = u(\boldsymbol{r}_A^*, \boldsymbol{r}_B, \boldsymbol{r}_C) - u(\boldsymbol{r}_A^0, \boldsymbol{r}_B, \boldsymbol{r}_C)$ and $h(\boldsymbol{r}_B, \boldsymbol{r}_C) = u(\boldsymbol{r}_C \mid \boldsymbol{r}_B^0, \boldsymbol{r}_C)$.

We now introduce a different class of independence concepts.

Definition 2.4.9. *Let $B_1, \ldots, B_k$ be a partition of $[m]$. Attributes $\boldsymbol{R}_{B_1}, \ldots, \boldsymbol{R}_{B_k}$ are said to be* additive independent *if the preference comparison of any two lotteries, defined as probability distributions over $\times_{i \in [k]} \boldsymbol{R}_{B_i}$, depends only on their marginal probability distributions.*

Additive independence can be generalised to the case where the indices of the attributes do not form a partition of $[m]$.

Definition 2.4.10. *Let $B_1, \ldots, B_k$ be such that $\cup_{i \in [k]} B_i = [m]$. Attributes $\boldsymbol{R}_{B_1}, \ldots, \boldsymbol{R}_{B_k}$ are said to be* **Generalized Additive Independent (GAI)** *if the preference comparison of any two lotteries depends only on the marginal probability distributions over $\times_{i \in [k]} \boldsymbol{R}_{B_i}$.*

2.4.2 Utility Factorisations

The sets of independences introduced in the previous section give rise to specific factorisations of the overall utility function. Suppose each function $u_A(\boldsymbol{r}_A)$, a marginal utility function over $\boldsymbol{R}_A$ only, is such that $u_i(\boldsymbol{r}_A^0) = 0$ and $u_i(\boldsymbol{r}_A^*) = 1$, $A \subseteq [m]$. The following results, from e.g. Keeney and Raiffa [1976], hold.

Proposition 2.4.11. *If a utility function $u(\boldsymbol{r})$ has m singly UI attributes, then it must take the form*

$$u(\boldsymbol{r}) = \sum_{A \in \mathcal{P}_0([m])} k_A \prod_{i \in A} u_i(r_i), \tag{2.4.2}$$

where $\mathcal{P}_0$ denotes the power set without the empty set and $k_A \in [0, 1]$ is a criterion weight [Keeney and Raiffa, 1976] such that $\sum_{A \subseteq [m]} k_A = 1$.

Definition 2.4.12. *Utility functions entertaining the factorisation in equation (2.4.2) are called* **multilinear***.*

The criterion weights of equation (2.4.2) can be evaluated by comparing the utility of terms $u(\boldsymbol{r}_A^*, \boldsymbol{r}_{[m] \setminus A}^0)$. Their actual form is not fundamental for this thesis and can be found in Keeney and Raiffa [1976].

Proposition 2.4.13. *If a utility function $u(\boldsymbol{r})$ has m mutually UI attributes, then it must take the form*

$$u(\boldsymbol{r}) = \sum_{A \in \mathcal{P}_0([m])} h^{n_A - 1} \prod_{i \in A} k_i u_i(r_i), \tag{2.4.3}$$

where n_A is the number of elements in A, $k_i = u(r_i^*, \boldsymbol{r}^0_{[n]\setminus\{i\}})$ and h is the unique solution not smaller than minus one to

$$1 + h = \prod_{i\in[m]} (1 + hk_i). \tag{2.4.4}$$

Definition 2.4.14. *Utility functions entertaining the factorisation in equations (2.4.3) and (2.4.4) are called* **multiplicative**.

Example 2.4.15. Consider three attributes R_1, R_2 and R_3. A multilinear factorisation over these attributes can be written as

$$u = k_1 u_1 + k_2 u_2 + k_3 u_3 + k_{12} u_1 u_2 + k_{13} u_1 u_3 + k_{23} u_2 u_3 + k_{123} u_1 u_2 u_3,$$

whilst the multiplicative one has the form

$$u = k_1 u_1 + k_2 u_2 + k_3 u_3 + h k_1 k_2 u_1 u_2 + h k_1 k_3 u_1 u_3 + h k_2 k_3 u_2 u_3 + h^2 k_1 h_2 k_3 u_1 u_2 u_3, \tag{2.4.5}$$

where for ease of notation we left the arguments of these functions implicit.

We can note that the multiplicative factorisation is a special case of the multilinear one. Importantly, in the multilinear case there are $2^m - 2$ criterion weights to elicit, whilst in the multiplicative one these are only m. Therefore the elicitation task is much simpler in the multiplicative case.

Proposition 2.4.16. *If a utility function $u(\boldsymbol{r})$ has m additive independent attributes then it must take the form*

$$u(\boldsymbol{r}) = \sum_{i\in[m]} k_i u_i(r_i), \tag{2.4.6}$$

where $\sum_{i\in[m]} k_i = 1$.

Definition 2.4.17. *Utility functions entertaining the factorisation in equation (2.4.6) are called* **additive**.

Additive utility functions can be considered as special cases of multiplicative utility functions, and therefore also of multilinear ones, by noting equation (2.4.6) coincides with equation (2.4.3) in the case the weights of Proposition 2.4.13 sum to unity.

Example 2.4.18. An additive factorisation over three attributes R_1, R_2 and R_3 can be written as

$$u(r_1, r_2, r_3) = k_1 u_1(r_1) + k_2 u_2(r_2) + k_3 u_3(r_3)$$

Proposition 2.4.19. *Suppose $\boldsymbol{R}_{B_1}, \ldots, \boldsymbol{R}_{B_k}$ are GAI, where $B_1, \ldots, B_k$ are such that $\cup_{i \in [k]} B_i = [m]$, then*

$$u(\boldsymbol{r}) = \sum_{i \in [k]} u_i(\boldsymbol{r}_{B_i}). \tag{2.4.7}$$

Example 2.4.20. Consider again three attributes R_1, R_2 and R_3 and assume $\{R_1, R_2\}$ and $\{R_2, R_3\}$ are GAI. Then

$$u(r_1, r_2, r_3) = u_{12}(r_1, r_2) + u_{23}(r_2, r_3). \tag{2.4.8}$$

Note however that the functions u_{12} and u_{23} are not uniquely defined. Braziunas and Boutilier [2005] and Fishburn [1967] showed that in this example the utility function in general decomposes as

$$u(r_1, r_2, r_3) = u(r_1, r_2, r_3^0) + u(r_1^0, r_2, r_3) + u(r_1^0, r_2, r_3^0). \tag{2.4.9}$$

Therefore the third term on the right hand side (rhs) of equation (2.4.9) can be associated to either of the first two terms of the rhs of equation (2.4.8). Fishburn [1967] defined the so called canonical decomposition which uniquely identifies the subutility functions.

2.4.3 Graphical Utility Models

The most general case of the previous section still assumes that all the attributes are utility independent of their complement. We note here that there are situations where this assumption is not tenable and at least one attribute is not utility independent of its complement. Such a situation is usually referred to as *partial utility independence*. Abbas [2010] derived a general expansion theorem for multiattribute utility functions which decomposes the overall function into a linear combination of products of conditional utility functions of Definition 2.4.7. The resulting expression can then be simplified using CUI statements similarly to conditional independence in probabilistic reasoning. The underlying CUI structure can be represented by a network, exhibiting the same expressive power of network representations of probabilistic conditional independences.

For the purpose of this thesis we mostly focus on the work of Abbas [2010], but we note here that other authors have proposed solutions to factorise multiattribute utility functions in the partial utility independence case. Farquhar [1975] presented a decomposition theorem for utility independence structures called fractional hypercubes; Abbas and Howard [2005] introduced a class of utility functions called attribute dominance utility, where utility is equal to zero whenever any attribute is at its worst outcome, and deduced expansion theorems for such a class; Bell [1979] introduced multiattribute factorisations using interpolation techniques.

We also note that a variety of graphical representations of different sets of preferential independences have been defined. Again in this thesis we mostly focus on Abbas [2010] which introduced *bidirectional utility diagrams* to represent sets of CUI statements. These diagrams are a generalisation of the utility diagrams of Abbas and Howard [2005], which only describe attribute dominance utilities. In the following chapters we restrict our attention to a particular subclass of bidirectional utility diagrams, that can also be thought of as CUI networks [Engel and Wellman, 2008]. Graphical models based on additive independences also exist [Bacchus and Grove, 1995, Braziunas and Boutilier, 2005]. Lastly, we note that La Mura and Shoham [1999] introduced expected utility networks that simultaneously represent probabilistic and preferential independences.

We first present the expansion theorem of Abbas [2010]. Let $\boldsymbol{\mathcal{R}}_A^{0*}$ be the set of all possible instantiations of $\boldsymbol{R}_A$ such that, for $i \in A$, R_i is either at its maximum or at its minimum value, $A \subset [m]$. Suppose A is totally ordered and, for $i \in A$, let B_i be the subset of A including the successors of i in A according to this order. Further let $V_i = A \setminus \{B_i \cup \{i\}\}$.

Proposition 2.4.21. *For any $A \subseteq [m]$, $u(\boldsymbol{r})$ can be written as*

$$u(\boldsymbol{r}) = \sum_{\boldsymbol{r}_A^{0*} \in \boldsymbol{\mathcal{R}}_A^{0*}} u(\boldsymbol{r}_A^{0*}, \boldsymbol{r}_{[m]\setminus A}) \prod_{i \in A} g(r_i \mid \boldsymbol{r}_{B_i}^{0*}, \boldsymbol{r}_{V_i}), \qquad (2.4.10)$$

where

$$g(r_i \mid \boldsymbol{r}_{B_i}^{0*}, \boldsymbol{r}_{V_i}) = \begin{cases} u(r_i \mid \boldsymbol{r}_{B_i}^{0*}, \boldsymbol{r}_{V_i}), & \text{if } r_i = r_i^* \text{ in } u(\boldsymbol{r}_A^{0*}, \boldsymbol{r}_{[m]\setminus A}), \\ \check{u}(r_i \mid \boldsymbol{r}_{B_i}^{0*}, \boldsymbol{r}_{V_i}), & \text{if } r_i = r_i^0 \text{ in } u(\boldsymbol{r}_A^{0*}, \boldsymbol{r}_{[m]\setminus A}). \end{cases}$$

Example 2.4.22. Consider three attributes R_1, R_2 and R_3. A generic decomposition for $u(r_1, r_2, r_3)$ assuming $3 \succ 2 \succ 1$, where $\succ$ is the order relation, can be written as

$$
\begin{aligned}
u(r_1, r_2, r_3) =\ & u(r_1^*, r_2^*, r_3^*)u(r_3 \mid r_2, r_1)u(r_2 \mid r_3^*, r_1)u(r_1 \mid r_3^*, r_2^*) + \\
& u(r_1^0, r_2^*, r_3^*)u(r_3 \mid r_2, r_1)u(r_2 \mid r_3^*, r_1)\check{u}(r_1 \mid r_3^*, r_2^*) + \\
& u(r_1^*, r_2^0, r_3^*)u(r_3 \mid r_2, r_1)\check{u}(r_2 \mid r_3^*, r_1)u(r_1 \mid r_3^*, r_2^0) + \\
& u(r_1^*, r_2^*, r_3^0)\check{u}(r_3 \mid r_2, r_1)u(r_2 \mid r_3^0, r_1)u(r_1 \mid r_3^0, r_2^*) + \\
& u(r_1^0, r_2^0, r_3^*)u(r_3 \mid r_2, r_1)\check{u}(r_2 \mid r_3^*, r_1)\check{u}(r_1 \mid r_3^*, r_2^0) + \\
& u(r_1^0, r_2^*, r_3^0)\check{u}(r_3 \mid r_2, r_1)u(r_2 \mid r_3^0, r_1)\check{u}(r_1 \mid r_3^0, r_2^*) + \\
& u(r_1^*, r_2^0, r_3^0)\check{u}(r_3 \mid r_2, r_1)\check{u}(r_2 \mid r_3^0, r_1)u(r_1 \mid r_3^0, r_2^0). \qquad (2.4.11)
\end{aligned}
$$

The first term in each monomial can be written as a linear combination of criterion weights, whilst all the other conditional utilities have a single argument.

FIGURE 2.17: Example of a bidirectional utility diagram over three attributes.

The utility decomposition in equation (2.4.10) can be simplified using various sets of CUIs. These can be represented by the so called bidirectional utility diagrams. Here we slightly change the definition of Abbas [2010] and represent these as CGs.

Definition 2.4.23. *A* bidirectional utility diagram *is a chain graph $\mathcal{G}$ with vertex set $\{R_1, \ldots, R_m\}$ and whose edges represent the possibility of utility dependence as follows:*

- *if $(R_i, R_j), (R_j, R_i) \notin E(\mathcal{G})$, then R_i is CUI of R_j given $\boldsymbol{R}_{[m]\setminus\{i,j\}}$ and R_j is CUI of R_i given $\boldsymbol{R}_{[m]\setminus\{i,j\}}$;*

- *if $(R_i, R_j) \in E(\mathcal{G})$ but $(R_j, R_i) \notin E(\mathcal{G})$ then R_i is CUI of R_j given $\boldsymbol{R}_{[m]\setminus\{i,j\}}$;*

- *if $(R_i, R_j), (R_j, R_i) \in E(\mathcal{G})$ then no utility independence is implied.*

Example 2.4.24. In Figure 2.17 we present an example of a bidirectional diagram. This diagram includes only a directed edge and therefore represents the closest situation to a multilinear factorisation, corresponding to a diagram with no edges. This diagram asserts that R_1 is UI of $\{R_2, R_3\}$ and vice versa, plus that R_3 is CUI of R_2 given R_1. Therefore, using the CUI statements associated to the diagram in Figure 2.17, equation (2.4.11) can be rewritten as

$$
\begin{aligned}
u &= u(r_1^*, r_2^*, r_3^*)u(r_3 \mid r_2^*)u(r_2 \mid r_3^*)u(r_1) + u(r_1^0, r_2^*, r_3^*)u(r_3 \mid r_2^*)u(r_2 \mid r_3^*)\breve{u}(r_1) + \\
&\quad u(r_1^*, r_2^0, r_3^*)u(r_3 \mid r_2^0)\breve{u}(r_2 \mid r_3^*)u(r_1) + u(r_1^*, r_2^*, r_3^0)\breve{u}(r_3 \mid r_2^0)u(r_2 \mid r_3^0)u(r_1) + \\
&\quad u(r_1^0, r_2^0, r_3^*)u(r_3 \mid r_2^0)\breve{u}(r_2 \mid r_3^*)\breve{u}(r_1) + u(r_1^0, r_2^*, r_3^0)\breve{u}(r_3 \mid r_2^*)u(r_2 \mid r_3^0)\breve{u}(r_1) + \\
&\quad u(r_1^*, r_2^0, r_3^0)\breve{u}(r_3 \mid r_2^0)\breve{u}(r_2 \mid r_3^0)u(r_1).
\end{aligned}
$$

Abbas [2010] identifies classes of bidirectional utility diagrams which give rise to optimal expansions of the multiattribute utility function, i.e. those where the conditional utility functions include the lowest number of arguments.

For the purpose of this thesis we consider only a subclass of bidirectional utility diagrams. This was defined in Leonelli and Smith [2015].

Definition 2.4.25. *We say that a bidirectional utility diagram is a* directional utility diagram *if its graph is a DAG.*

Note that the diagram in Figure 2.17 is unconventional in the sense that an attribute is parent of another one having a lower index. We motivate this choice in Chapter 4,

where such diagrams proves to be very useful. Directional utility diagrams are extremely powerful in the multi-expert domains we address in later chapters, since the utility function associated to such diagrams can be written in terms of criterion weights and univariate utility functions. This is formalised in the following lemma.

Lemma 2.4.26. *For a directional utility diagram with vertex set $\{R_i : i \in [m]\}$ there exists an expansion order such that equation (2.4.10) is a linear combination of terms involving only criterion weights and conditional utility/disutility function having as argument a single attribute.*

Proof. This result follows by observing that the terms $u(\boldsymbol{r}_A^{0*}, \boldsymbol{r}_{[m]\setminus A})$ in equation (2.4.10) coincide with $u(\boldsymbol{r}_{[m]}^{0*})$ since the expansion can be performed over all the attributes. These terms are functions of criterion weights. Furthermore the conditional independence structure underlying a directed utility diagram is such that there is an expansion order where $R_i \perp\!\!\!\perp \boldsymbol{R}_{V_i} \mid \boldsymbol{R}_{[m]\setminus\{V_i\cup i\}}$. Thus $g(r_i \mid \boldsymbol{r}_{B_i}^{0*}, \boldsymbol{r}_{V_i})$ in equation (2.4.10) is equal to $g(r_i \mid \boldsymbol{r}_{B_i}^{0*})$. $\square$

2.4.4 Univariate Utility Theory

In the previous sections we have shown how multiattribute utility problems can be decomposed into univariate ones when various sets of independences are believed to hold. We now discuss some of the characteristics of univariate utility functions and provide various examples.

First, recall that utility functions are unique up to affine transformations, meaning that they imply the same preference ranking. Two such functions are said to be *strategically equivalent*. In addition, Keeney and Raiffa [1976] showed that if two utility functions are strategically equivalent, then one must be an affine transformation of the other.

We now introduce a variety of utility functions widely used in practice.

Example 2.4.27. For the purpose of this example, we make explicit the dependence on the decision $\boldsymbol{d}$. Choices for $u_i(r_i, \boldsymbol{d})$ are

- the *linear utility* function, $u_i(r_i, \boldsymbol{d}) = a(\boldsymbol{d}) + b(\boldsymbol{d})r_i$, for $b(\boldsymbol{d}) > 0$;

- the *quadratic utility*, $u_i(\boldsymbol{d}, r_i) = 1 - (\rho_i(\boldsymbol{d})r_i)^2$, whose parameter is specified through the ideal scenario $\rho_i(d)$;

- the *polynomial utility* of degree l, $u_i(\boldsymbol{d}, r_i) = \sum_{j\in[l]} \rho_{ij}(\boldsymbol{d})r_i^j$, where both the coefficients $\rho_{i,j}(\boldsymbol{d})$ and the domain of the rewards need to entertain some constraints

[see Keeney and Raiffa, 1976, Müller and Machina, 1987, for a discussion of these, which we omit here since are not relevant for our developments];[7]

- the *exponential utility* function, $u_i(r_i, \boldsymbol{d}) = 1 + \exp\{-\rho_i(\boldsymbol{d})r_i\}$, where the parameter $\rho_i(\boldsymbol{d}) > 0$, reflects the degree of risk aversion.

Many of the characteristics of a univariate utility function, as for example whether it is monotonic and convex or not, describe the preferences of the DM and her attitude towards risk. See for more details on this Keeney and Raiffa [1976].

2.5 Graphical Decision Models

In the previous sections we assumed the existence of a generic decision space $\boldsymbol{\mathcal{D}}$ representing the available choices to a DM. However, in practice, DMs have to commit to a sequence of decisions, where some of the consequences of committing to a certain action are observed before having to choose subsequent policies. Such a sequential structure can be depicted through graphical representations of various types. Here we focus on the Influence Diagram (ID) model, which can be thought of as an augmented version of a BN including also decision and utility vertices. These are introduced in Section 2.5.1. We note that, although widely used in practice, IDs are able to represent decision problems entertaining fierce symmetric conditions. In Section 2.5.2 we provide a brief overview of the available models for the so called asymmetric decision problems.

For the purpose of this section, we slightly change our notation, which turns out to be very useful in concisely represent many of the features related to IDs. Furthermore, as shown in Section 5.6.1, this enables a straightforward implementation of the study of these models in a computer algebra system. This is a new representation of IDs that we introduced in Leonelli et al. [2015a]. Let $[n]$ be partitioned in $\mathbb{V}$ and $\mathbb{D}$, where $\mathbb{V}$ is the index set of the random (or non-controlled) variables, whilst $\mathbb{D}$ is the index set of the decisions (or controlled variables).

Example 2.5.1. Let $n = 6$ and assume $\mathbb{D} = \{1, 4\}$ and $\mathbb{V} = \{2, 3, 5, 6\}$. Then Y_1 and Y_4 are decisions, whilst Y_2, Y_3, Y_5 and Y_6 are random variables.

2.5.1 Influence Diagrams

As noted above, only decision problems enjoying strong symmetry conditions can be represented as IDs [Howard and Matheson, 1983, 2005b, Jensen and Nielsen, 2009, Shachter,

---

[7]Note that the polynomial utility has no monomial of degree zero. This is because many of the results of Chapter 5 are more easily interpretable in this case. Recall that adding such a monomial does not change the analysis as the two utilities are strategically equivalent.

1986, Shachter and Kenley, 1989, Smith, 2010]. We define here, following [Smith, 2010, Smith and Thwaites, 2008], the class of problems that entertains such a representation, that we call uniform.

Definition 2.5.2. *A* uniform decision problem *is such that:*

- *a variable Y_i, $i \in \mathbb{D}$, is controlled before controlling Y_j, $j \in \mathbb{D}$, if $i < j$, and is remembered when making decisions with a higher index;*[8]

- *the space $\boldsymbol{\mathcal{Y}}_{[n]}$ is the product of the individual spaces, i.e. $\boldsymbol{\mathcal{Y}}_{[n]} = \times_{i \in [n]} \mathcal{Y}_i$;*

- *if a random variable Y_i, $i \in \mathbb{V}$, is observed before a variable Y_j, then $i < j$ and its distribution can depend on $\boldsymbol{Y}_{\mathbb{D}}$ only through $\boldsymbol{Y}_{\mathbb{D} \cap [i-1]}$.*

The condition in the first bullet is usually referred to as no-forgetting. This bullet further implies that decisions are totally ordered. Furthermore decisions and random variables cannot be constrained with each other in uniform decision problems, as specified by the second bullet. The third bullet corresponds to the causal consistency lemma of Cowell et al. [1999] and guarantees that decisions cannot influence variables that have been already observed.

As briefly mentioned, the vertex set of an ID consists of random variables, decisions and utilities. We focus now on the latter type of nodes. The ID literature usually assumes that there is an additive factorisation between the utilities associated to different vertices. We refer to such IDs as *additive IDs*. In Leonelli et al. [2015a] we defined the **Multiplicative Influence Diagram (MID)** which considers the much larger class of multiplicative utility factorisations. Let $\mathbb{U} = \{u_i : i \in [m]\}$ and u_i, $i \in [m]$, maps a subspace $\boldsymbol{\mathcal{Y}}_{P_i}$ of $\boldsymbol{\mathcal{Y}}_{[n]}$ into $[0, 1]$, with $P_i \subseteq [n]$.

Definition 2.5.3. *An* MID $\mathcal{G}$ *consists of*

- *a DAG with vertex set $V(\mathcal{G}) = \{Y_i, u_j : i \in [n], j \in [m]\}$ and edge set $E(\mathcal{G})$ including three types of edges defined by the following rules:*

  - *for $i \in [m]$, u_i has no children. Its parent set is $\{Y_j : j \in P_i\}$ and an index $i \in [n]$ can appear in a set P_j only, $j \in [m]$, i.e. an element of $\{Y_i : i \in [n]\}$ is parent of at most one element of $\mathbb{U}$. For u_i and u_j, $i, j \in [m]$, $i > j$ if there is a $k \in P_i$ such that for every $l \in P_j$, $k > l$: i.e. a utility node u_i has higher index than u_j if one parent of u_i has higher index than all parents of u_j;*

  - *for $i \in \mathbb{D}$, the parent set of Y_i, $\{Y_j : j \in \Pi_i\}$, for $\Pi_i \subset [n]$, consists of the random variables and the decisions that are known when Y_i is controlled;*

---

[8]With the term *controlled* we mean that a decision variable is implemented.

          − *for $i \in \mathbb{V}$, the parent set of Y_i, $\{Y_j : j \in \Pi_i\}$, for $\Pi_i \subset [n]$, is such that $Y_i \perp\!\!\!\perp \boldsymbol{Y}^i_{[n]} \mid \boldsymbol{Y}_{\Pi_i}, \boldsymbol{\theta}$;*

- *for $i \in \mathbb{V}$, a density function $f_i(y_i \mid \boldsymbol{y}_{\Pi_i}, \boldsymbol{\theta}_i)$;*

- *a multiplicative utility factorisation over $\mathbb{U}$.*

Therefore, in an MID there are three types of vertices and three types of edges. The vertices are either random variables, decisions or utilities. Random variables are framed in circles, decisions in squares, whilst utilities have no framing in the DAG $\mathcal{G}$ of an MID. An edge into a vertex associated to a random variable asserts that the density of that random variable is conditional on the variable associated to the parent vertex. This edge therefore represents probabilistic dependence. Edges into decision vertices are called *information* edges. These imply that the variables in the parent vertex are known when making that decision. Edges into utility vertices simply represent functional relationships, meaning that the utility is a function of the parent variables. Smith [2010] defined the reduced ID, where some of the information edges are not included, thus rendering the overall graphical representation clearer. Specifically, in a reduced ID no edge (Y_k, Y_i) is included if $(Y_k, Y_j) \in E(\mathcal{G})$, $k < j < i$, $j, i \in \mathbb{D}$.

IDs are extremely powerful graphical representations of the qualitative structure underlying the elements of decision problems. In particular, these are able to depict both continuous and discrete domains. Note that although P_i, $i \in [m]$, is effectively an index parent set, we had to use a different notation than the one generically introduced for graphs in Appendix B, because of the existence of both utility and non-utility nodes.

Definition 2.5.3 implies a total order over the vertex set of an MID. For $i \in [m]$, let j_i be the highest index of P_i and $\mathbb{J} = \{j_i : i \in [m]\}$. The totally ordered sequence associated to $V(\mathcal{G})$ is called **Decision Sequence (DS)** of the MID $\mathcal{G}$ and indicated as $S := (Y_1, \ldots, Y_{j_1}, u_1, Y_{j_1+1}, \ldots, Y_{j_m}, u_m)$. In contrast with other authors [e.g. Cowell et al., 1999, Jensen and Nielsen, 2009], we do not introduce utility nodes only at the end of the DS. This entails an efficient computation of the expected utility scores, since optimal decisions can be shown to be functions of smaller sets of preceding variables.

Example 2.5.4. Figure 2.18 presents an MID with vertex set $\{Y_1, \ldots, Y_6, u_1, \ldots, u_3\}$ describing a gross simplification of the possible countermeasures after an accident at a nuclear power plant. For this MID, $n = 6$, $m = 3$, $\mathbb{D} = \{1, 4\}$ and $\mathbb{V} = \{2, 3, 5, 6\}$. Therefore, there are two decisions Y_1 and Y_4: the first consisting of the possibility of closing down the power plant, the second of evacuating the nearby population. Before controlling Y_4, the variables Y_1, Y_2 and Y_3 are observed since $\Pi_4 = \{1, 2, 3\}$. There are four random nodes: Y_2 and Y_3 measure the amount of dispersion of the contaminant
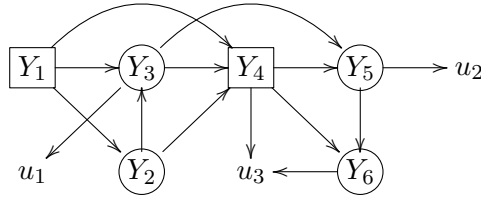
FIGURE 2.18: An influence diagram describing the available countermeasures after an accident at a nuclear power plant.

in the atmosphere and to the ground respectively, Y_5 estimates the human intake of radiation and Y_6 ranks the level of stress in the population. This DAG implies for example that $Y_5 \perp\!\!\!\perp Y_2 \mid Y_3$ for every $y_1 \in \mathcal{Y}_1$ and $y_4 \in \mathcal{Y}_4$. The variable Y_5 has parent set $\{Y_3, Y_4\}$, $\Pi_5 = \{3, 4\}$ and $\Pi_{5 \cap \mathbb{V}} = \{3\}$. We further assume all variables to be binary, taking values in the spaces $\mathcal{Y}_i = [1]^0$, $i \in [6]$. When $\mathcal{Y}_i$ is associated to a decision node, 1 and 0 correspond to, respectively, proceed and not proceed. If $\mathcal{Y}_i$ is associated to a random node, then 1 and 0 correspond respectively to high and low. The MID definition is completed by three utility nodes, u_1, u_2 and u_3. For example the set P_3 is equal to $\{4, 6\}$ and therefore Y_4 and Y_6 are the arguments of the utility function u_3. Lastly, the DS $(Y_1, Y_2, Y_3, u_1, Y_4, Y_5, u_2, Y_6, u_3)$ is the one associated to the MID in Figure 2.18, where $j_1 = 3$, $j_2 = 5$, $j_3 = 6$ and $\mathbb{J} = \{3, 5, 6\}$.

The density associated to an MID enjoys a factorisation which mirrors the one of BNs.

Proposition 2.5.5. *The probability density function associated to an MID can be written as* $f(\boldsymbol{y}_\mathbb{V} \mid \boldsymbol{\theta}) = \prod_{i \in \mathbb{V}} f(y_i \mid \boldsymbol{y}_{\Pi_i}, \boldsymbol{\theta}_i)$, *where* $\boldsymbol{\theta}^\mathrm{T} = (\boldsymbol{\theta}_i)_{i \in \mathbb{V}}$.

Example 2.5.6. The density function associated to the MID in Figure 2.18 can be written as

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = f(y_2 \mid y_1, \boldsymbol{\theta}_2) f(y_3 \mid y_2, y_1, \boldsymbol{\theta}_3) f(y_5 \mid y_4, y_3, \boldsymbol{\theta}_5) f(y_6 \mid y_5, y_4, \boldsymbol{\theta}_6).$$

The utility factorisation associated to this MID coincides with the one in equation (2.4.5)

2.5.1.1 The Evaluation of Influence Diagrams. To *evaluate* an MID is to identify a sequence of optimal decisions maximising the expected utility function. However, only MIDs whose topology is such that, for any index $j \in \mathbb{D}$, the only variables that are known at the time the DM makes a decision Y_j have an index lower than j, can be directly evaluated using standard 'extensive form' calculations [Smith, 1989b]. This is because the evaluation outputs optimal decisions as functions of observed quantities only and therefore DMs can unambiguously commit to such an optimal course of action. Here we define this class using the terminology of Smith [2010].

Definition 2.5.7. *An MID* $\mathcal{G}$ *is in* extensive form *if* $i \in \Pi_j$, *for all* $j \in \mathbb{D}$ *and* $i \in [j-1]$.

We assume here for the sake of simplicity that the MID under study is in extensive form. However, we note here that any MID in non extensive form can be transformed into one that is: albeit sometimes with the loss of some efficiency. There are several rules, which we do not introduce here, to manipulate the topology of an MID, as for example, edge reversal and barren node removal [see e.g. Jensen and Nielsen, 2009, Shachter, 1986], such that the resulting MID is in extensive form. Without restriction we also assume that every vertex Y_i of the MID, $i \in [n]$, has at least one child. Random and controlled vertices with no children could be simply deleted from the graph without changing the outcome of the evaluation of the MID [see e.g. Jensen and Nielsen, 2009].

Example 2.5.8. The MID in Figure 2.18 is in extensive form since $\Pi_4 = \{1, 2, 3\}$. If either the edge (Y_2, Y_4) or (Y_3, Y_4) were deleted from this diagram, then the MID would not be in extensive form any more. Furthermore note that the only vertices with no children are u_1, u_2 and u_3.

Several algorithms have now been defined to evaluate IDs. These can work directly on the ID [Olmsted, 1984, Shachter, 1986], transform it into a decision tree [Canbolat et al., 2007], or convert the diagram into some other graphical structure [Jensen et al., 1994, Madsen and Jensen, 1999]. There were severe computational issues related to the evaluation of IDs, but there are now several promising methodologies designed to overcome these challenges, as noted in Bielza et al. [2011] and Gómez et al. [2007]. Many different pieces of software are also now available to build and automatically evaluate large IDs [Jensen and Nielsen, 2013]. The simplest way to evaluate an MID in extensive form is through a backward inductive algorithm over the vertices of the DAG [see e.g. Jensen and Nielsen, 2009]. Here we introduce a computationally efficient new version of this algorithm, including at each step only the strictly necessary utility nodes. For simplicity, we work with the predictive distribution $f(\boldsymbol{y}) = \int_{\boldsymbol{\Theta}} f(\boldsymbol{y} \mid \boldsymbol{\theta}) f(\boldsymbol{\theta})$, but the result can be straightforwardly generalised to take into account the parameter vector. The identification of the optimal policy is based on the computation of the functions $\bar{u}_i(\boldsymbol{y}_{B_i})$, $i \in [n]$, we formally introduce in Proposition 2.5.10, where

$$B_i = \left\{ \bigcup_{k \in [n]_{i-1} \cap \mathbb{V}} \Pi_k \quad \bigcup_{j \in [n]_{i-1} \cap \mathbb{J}} P_j \right\} \setminus [n]_{i-1}, \qquad (2.5.1)$$

is the index set of the variables that appear as arguments of $\bar{u}_i$, where $[n]_{i-1} = \{i, \ldots, n\}$. It should be noted that B_i includes only indices smaller than i that are either in the parent set of a random variable Y_k, $k > i$, or in a set P_j such that u_j succeeds Y_i in the DS of the MID.

Example 2.5.9. We compute here the sets B_5 and B_4 associated with the MID in Figure 2.18. The set $B_5 = \{3, 4\}$ since $B_5 = \{\Pi_6 \cup \Pi_5 \cup P_3 \cup P_2\} \setminus \{5, 6\}$, where $\Pi_6 = \{4, 5\}$,

$\Pi_5 = \{3, 4\}$, $P_3 = \{4, 6\}$ and $P_2 = \{5\}$. Furthermore $B_4 = \{3\}$ since it can be noted that $B_4 = B_5 \setminus \{4\}$.

Proposition 2.5.10. *The optimal decision associated to an MID yields expected utility* $\bar{u}_1(\boldsymbol{y}_{B_1})$, *where, for* $i \in [n]$, $\bar{u}_i(\boldsymbol{y}_{B_i})$ *is defined as*

$$\bar{u}_i(\boldsymbol{y}_{B_i}) = \begin{cases} \bar{u}_i^{\mathbb{D}}(\boldsymbol{y}_{B_i}), & \text{if } i \in \mathbb{D}, \\ \bar{u}_i^{\mathbb{V}}(\boldsymbol{y}_{B_i}), & \text{if } i \in \mathbb{V}, \end{cases}$$

where, for $i = n$,

$$\bar{u}_n^{\mathbb{D}}(\boldsymbol{y}_{B_n}) = \max_{\mathcal{Y}_n} k_m u_m(\boldsymbol{y}_{P_m}), \qquad \bar{u}_n^{\mathbb{V}}(\boldsymbol{y}_{B_n}) = \int_{\mathcal{Y}_n} k_m u_m(\boldsymbol{y}_{P_m}) f_n(y_n \mid \boldsymbol{y}_{\Pi_n}) \mathrm{d}y_n, \quad (2.5.2)$$

and, for $i \in [n-1]$, *if* $i \in \mathbb{J}$ *and supposing* $i \in P_l$,

$$\bar{u}_i^{\mathbb{D}}(\boldsymbol{y}_{B_i}) = \max_{\mathcal{Y}_i} \Big(h k_l u_l(\boldsymbol{y}_{P_l}) \bar{u}_{i+1}(\boldsymbol{y}_{B_{i+1}}) + k_l u_l(\boldsymbol{y}_{P_l}) + \bar{u}_{i+1}(\boldsymbol{y}_{B_{i+1}}) \Big), \quad (2.5.3)$$

$$\bar{u}_i^{\mathbb{V}}(\boldsymbol{y}_{B_i}) = \int_{\mathcal{Y}_i} \Big(h k_l u_l(\boldsymbol{y}_{P_l}) \bar{u}_{i+1}(\boldsymbol{y}_{B_{i+1}}) + k_l u_l(\boldsymbol{y}_{P_l}) + \bar{u}_{i+1}(\boldsymbol{y}_{B_{i+1}}) \Big) f_i(y_i \mid \boldsymbol{y}_{\Pi_i}) \mathrm{d}y_i, \quad (2.5.4)$$

whilst, if $i \notin \mathbb{J}$,

$$\bar{u}_i^{\mathbb{D}}(\boldsymbol{y}_{B_i}) = \max_{\mathcal{Y}_i} \bar{u}_{i+1}(\boldsymbol{y}_{B_{i+1}}), \qquad \bar{u}_i^{\mathbb{V}}(\boldsymbol{y}_{B_i}) = \int_{\mathcal{Y}_i} \bar{u}_{i+1}(\boldsymbol{y}_{B_{i+1}}) f_i(y_i \mid \boldsymbol{y}_{\Pi_i}) \mathrm{d}y_n. \quad (2.5.5)$$

The proof of this result is given in Appendix A.1.1. In Section 5.6.1 we symbolically define MIDs and introduce a new symbolic evaluation algorithm.

2.5.2 Asymmetric Models

The ID framework is able to represent uniform decision problems only. However, real decision problems often exhibit asymmetries of various kinds. In Jensen and Nielsen [2009] asymmetries are categorised in three classes. If the possible outcomes or decision options of a variable vary depending on the past, the asymmetry is called *functional*. If the very occurrence of a variable depends on the past, the asymmetry is said to be *structural*. *Order* asymmetries are present if $\{Y_i : i \in \mathbb{D}\}$ is not totally ordered. We note here that there are four types of functional asymmetries that can occur:

- **chance** → **chance**: the outcome of random variables restricts the outcomes of other random variables;

- **chance** → **decision**: the outcome of random variables restricts the options of controlled variables;

- **decision** → **chance**: decisions taken restrict the possible outcomes of random variables;

- **decision** → **decision**: decisions taken restrict the options of other controlled variables.

Heuristically, for each of these asymmetries the observation of $\boldsymbol{y}_A$, $A \subset [n]$, changes the space $\boldsymbol{\mathcal{Y}}_B$ associated to a vector $\boldsymbol{Y}_B$, such that $A \cap B = \emptyset$. This new space, $\boldsymbol{\mathcal{Y}}'_B$ say, is a subspace of $\boldsymbol{\mathcal{Y}}_B$.

The simplest solution to the evaluation of an asymmetric problem is by embedding it in a symmetric one which can be represented as an ID. The above mentioned evaluation algorithms can then be used to identify optimal solutions in the embedded version of the problem. Clearly this methodology has several drawbacks. Most importantly it does not graphically represent the relevant asymmetries and often considers a much larger problem space than the original one. For this reason a variety of graphical models have been introduced to represent a variety of asymmetric constraints [Bhattacharjya and Shachter, 2012, Demirer and Shenoy, 2006, Jensen and Vomlelová, 2002, Jensen et al., 2006, Nielsen and Jensen, 2003, Smith et al., 1993]. It is not the purpose of this thesis to review these models and we briefly illustrate part of the semantic of the sequential ID [Jensen et al., 2006] model in the following example.

Example 2.5.11. Consider the ID in Figure 2.18 and assume that the DM now believes her decision problem is characterised by the following three asymmetries:

- whenever she decides to close the power source, then the population cannot be evacuated from the area: $Y_1 = 1 \Rightarrow Y_4 = 0$;

- if either the deposition or the dispersion levels in the area are high, then the human intake is high: $Y_2 = 1 \vee Y_3 = 1 \Rightarrow Y_5 = 1$;

- if the evacuation option is followed then both the human intake and the stress levels in the population are high: $Y_4 = 1 \Rightarrow Y_5 = 1 \wedge Y_6 = 1$.

Our graphical representation of these asymmetries is given in Figure 2.19, corresponding to a sequential ID [Jensen et al., 2006]. Asymmetries are represented as labels on additional dashed arcs. If the asymmetry is composite, meaning that it involves more than two variables, then vertices can be grouped through a dashed ellipse and dashed arcs can either start or finish by the side of these ellipses. As exemplified by the diagram in Figure 2.19, the graphical representation of asymmetric decision problems, although providing a framework to quickly identify optimal policies, often lose the intuitiveness
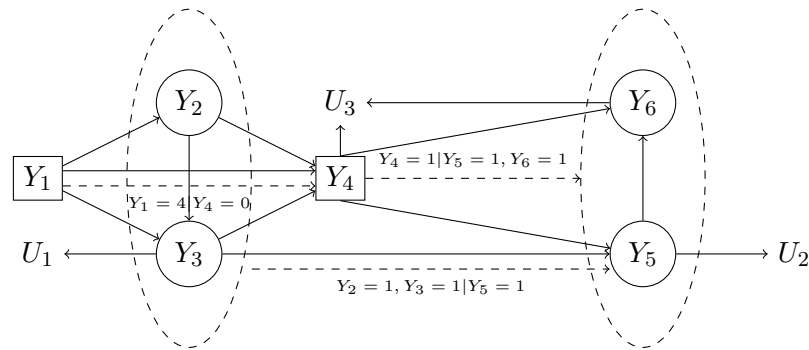
FIGURE 2.19: Representation of the asymmetric version of the multiplicative influence diagram of Figure 2.18 as a sequential influence diagram.

and the clarity of representation of standard IDs. We discuss more extensively this issue in Section 5.6.1.

Asymmetric decision problems can also be explicitly modelled as *decision trees* [see e.g. Clemen, 1996, French et al., 2009], an augmented version of an event tree including decision vertices and whose leaves are utilities. These suffer the same drawbacks of staged and event trees. Recently Cowell et al. [2010] introduced the decision CEG which can more compactly represent asymmetric decision problems.

2.6 Group Decision Making

The SEU model characterises rational decision making for single DMs. However, in practice, most decisions in organisations and society are the responsibility of groups rather than individuals: juries, cabinets and boards of directors are some examples. In addition, even single DMs rarely commit to certain courses of actions without seeking for the help of appropriate experts. How the suggestions of various experts have to be reconciled is a very interesting question and a variety of solutions have been proposed. Following French [2011], we identify two main contexts of group decision making:

- the **group decision problem**: a group of individuals is jointly responsible and accountable for the decision;

- the **expert problem**: a group of experts are consulted by a single DM who is responsible and accountable for the decision to be made.

French [2011] also identifies the *textbook problem* context which for the purpose of this thesis is not discussed here. Note further that a combination of the two above scenario can be considered in practice. There are two main categories of approaches for both contexts: behavioural and mathematical. We review the former ones in Section 2.6.1.

Mathematical approaches for the group decision problem are based on some algorithmic rule which, given the preferences and beliefs of each individual DM, then outputs the decision ranking of the group. We do not focus in this thesis on these methods and we simply note here that there is vast literature about these [see e.g. Arrow, 1951, French, 2011, French et al., 2009, Keeney, 1976, 2013, Keeney and Raiffa, 1976]. Most of these methods, embedding some idea of democratic voting, are subject to paradoxes and inconsistencies. We discuss mathematical approaches for the expert problem in Section 2.6.2.

2.6.1 Behavioural Approaches

Decision conferences and *facilitated workshops* are the most common techniques of behavioural aggregation of DMs' beliefs and preferences [Ackermann, 1996, Phillips, 1984, Phillips and Phillips, 1993]. These are usually two days events with around fifteen participants [although for example in the Chernobyl project we discussed in Chapter 1 the number of participants was way larger, as noted in French et al., 2009]. In such meetings, a *facilitator*, an individual skilled in the process of group discussion, encourages debate among the DMs and keeps the group's focus on the issue at hand. The facilitator is usually not an expert of the problem to be examined and has no ownership of the decision to be made. Each DM often has available interactive software to run her own analyses and observe the results of others' beliefs, which are also projected in the room. This software uses decision models which turns out to be fundamental to foster discussion and raise the main issues that need to be investigated by the group. The results of sensitivity analyses are also shown to the DMs in order to build consensus and avoid debates on irrelevant alternatives or parameters that do not affect the group's course of action.

As discussed in Chapter 1, these types of meetings were central to the decision elicitation process within the Chernobyl Project. The discussions at those meetings are summarised in French et al. [2009] and Smith [2010]. Ackermann [1996] reports more than 100 of successful applications of such behavioural techniques.

A different method called *Delphi protocol* has been designed to reach consensus by the remote and electronic compilation of a series of questionnaire, thus not requiring the group to meet [Dalkey and Helmer, 1963, Linstone and Turoff, 1975]. Under this protocol, each individual anonymously responds to the questionnaire. The answers are then sent to the other members for further comments and generations of possible courses of actions. This sequence is then iteratively repeated until consensus is reached.

2.6.2 The Expert Problem

The second context where groups are involved in the decision making process is the expert problem. Here an individual DM needs to synthesise the beliefs of a group of m experts into a unique probability distribution. Mathematical methods for the aggregation of expert judgement can be broadly categorised into two main branches: Bayesian and pooling operator methods. Before discussing these approaches, we remark that it is not the scope of this thesis to explore the validity of each of these [see e.g. Clemen and Winkler, 1999, French, 2011, and the references therein for an extensive discussion]. O'Hagan et al. [2006] discussed protocols to be followed by each expert to elicit his own probability distribution, whilst Merrick [2008] introduced methods to identify the right mix of experts to obtain the best aggregated belief. Here we mostly focus on pooling operators and we simply note that the Bayesian approach treats experts' judgements as data, synthesised in some appropriate likelihood function, to update the prior probability of the DM [see e.g. Albert et al., 2012, French, 1980, Mumpower and Stewart, 1996, Wiper and French, 1995].

2.6.2.1 Pooling Operators. Given the individual probabilities of a group of m experts, $f^i(\boldsymbol{y})$, $i \in [m]$, pooling operators consist of a function T^p which pools together into a unique probability distribution, $f^{DM}(\boldsymbol{y})$, the probability distributions of each member of the group. For ease of notation we suppress the dependence on the parameter vector.

The simplest operator is the *linear* one [Stone, 1961], corresponding to a weighted average of the experts' distributions:

$$f^{DM}(\boldsymbol{y}) = \mathrm{T}^p(f^1(\boldsymbol{y}), \ldots, f^m(\boldsymbol{y})) = \sum_{e \in [m]} w^e f^e(\boldsymbol{y}).$$

Bacharach [1975] introduced the *logarithmic Operator (logOp)* defined as

$$f^{DM}(\boldsymbol{y}) = \mathrm{T}^p(f^1(\boldsymbol{y}), \ldots, f^m(\boldsymbol{y})) \propto \prod_{e \in [m]} f^e(\boldsymbol{y})^{w^e}.$$

One of the main issues concerning pooling operators is the identification of the weights w_e to be given to each expert. The aggregation most commonly used in practice is through a linear operator whose weights are calibrated according to the forecasting performance of the experts on a calibration set of uncertain quantities. This method is usually called *classical* approach to the combination of expert judgements [Cooke, 1991].

Sets of axioms characterising operators encoding rational behaviour have been defined [see for a review Genest et al., 1986]. We argue here, following Faria and Smith [1997],

that one very desirable property for a pooling operator is to be *Externally Bayesian (EB)* [Madansky, 1964].

Definition 2.6.1. *A pooling operator* T^p *is said to be EB if* $f^{DM}(\boldsymbol{y})$ *is the same when the following two procedures are implemented:*

- *first apply* T^p *and then update* $f^{DM}(\boldsymbol{y})$ *using Bayes theorem;*

- *first update each expert's distribution* $f^i(\boldsymbol{y})$ *via Bayes Theorem and then apply* T^p.

Therefore for EB operators the order in which pooling and updating are performed is irrelevant. This is a desirable property for a variety of reasons: first, experts do not need to meet every time data is collected; second, experts cannot increase their influence on the outcome by choosing whether to include their beliefs before or after the collection of data; lastly, as we discuss below, operators exhibiting such a property are in general able to retain the conditional independences underlying an agreed model. Importantly, Genest [1984] characterised the logOp as the only operator being EB. Although EB operators have useful properties in the domains we are addressing, we note that these operators have been criticised in the literature [see e.g. Lindley, 1985].

2.6.2.2   The Multivariate Case. When experts have to elicit probabilities in large multivariate settings, their performance might dramatically decrease because of the difficulty in assessing correlations [Clemen et al., 2000, Winkler and Clemen, 2004]. One approach often used to mitigate such difficulties is to elicit probabilities over an agreed qualitative graphical structure, as for example a BN [Burns and Clemen, 1993, Faria and Smith, 1997, Farr et al., 2015, Renooij, 2001, Smith and Faria, 2000]. Smith [1996] extensively argued that experts' agreement can be more easily found in graphical frameworks since the underlying conditional independence structure can be discussed in natural language.

Pennock and Wellman [2005] proved that a pooled distribution obtained with a logOp reflects any shared Markov independence by the group and for this reason, it can be represented in a concise and natural manner with a BN as well. However, it was also noted that no pooling operator can maintain all independences representable in a BN. Having acknowledged this negative result, in Chapter 3 we develop a framework where groups of experts qualitatively agree on a specific graphical model and then quantitatively individually deliver probability judgements, preserving the group's conditional independence structure.

Here we focus on the work of Faria and Smith [1996] and Faria and Smith [1997] which assumes the presence of an underlying decomposable CG agreed by all the experts.

Their work, introducing a generalisation of the logOp, falls within the pooling operators approach. Suppose the decomposable CG $\mathcal{G}$ has n chain components $\{\boldsymbol{Y}_i : i \in [n]\}$.

Definition 2.6.2. *For a decomposable CG $\mathcal{G}$, the **conditional logOp**, T^P, is defined by the components $\mathrm{T}^p_{\Pi_j}$*

$$f_j^{DM}(\boldsymbol{y}_j \mid \boldsymbol{y}_{\Pi_j}) = \mathrm{T}^p_{\Pi_j}(\cdot) \propto \prod_{i \in [m]} f_j^i(\boldsymbol{y}_j \mid \boldsymbol{y}_{\Pi_j})^{w_j^i(\Pi_j)},$$

where f_j^i is is the conditional density relative to $\boldsymbol{Y}_j$ provided by the i-th expert and $w_j^i(\Pi_j)$ is a weight. The DM distribution is then defined as

$$f^{DM}(\boldsymbol{y}) = \mathrm{T}^p = \prod_{j \in [n]} \mathrm{T}^p_{\Pi_j} = \prod_{j \in [n]} f_j^{DM}(\boldsymbol{y}_j \mid \boldsymbol{y}_{\Pi_j}).$$

Although the conditional logOp is not EB, Faria and Smith [1997] showed that it respects a weaker condition called Conditional External Bayesianity (CEB).

Definition 2.6.3. *Given a CG $\mathcal{G}$, **CEB** holds if the conditional densities $f_j^i(\boldsymbol{y}_j \mid \boldsymbol{y}_{\Pi_j})$, $i \in [m]$, $j \in [n]$, are pooled with an EB operator.*

We define here a class of likelihood functions that guarantees CEB when the pooling is performed with a conditional logOp.

Definition 2.6.4. *A likelihood $l(\boldsymbol{x} \mid \boldsymbol{y})$ for a sample $\boldsymbol{x}^{\mathrm{T}} = (\boldsymbol{x}_i)_{i \in [n]}$ is in the class of cutting *likelihoods of a CG $\mathcal{G}$ with chain components $\{\boldsymbol{Y}_i : i \in [n]\}$ if*

$$l(\boldsymbol{x} \mid \boldsymbol{y}) = l_1(\boldsymbol{x}_1 \mid \boldsymbol{y}_1) l_2(\boldsymbol{x}_2 \mid \boldsymbol{y}_2, \boldsymbol{y}_{\Pi_2}, \boldsymbol{x}_1) \cdots l_n(\boldsymbol{x}_n \mid \boldsymbol{y}_n, \boldsymbol{y}_{\Pi_n}, \boldsymbol{x}^{n-1}). \qquad (2.6.1)$$

Faria and Smith [1997] then proved that the conditional logOp respects CEB when the pooling/updating is performed according to a procedure we briefly review here. Let $\tilde{f}_g$ be the joint density over $\mathcal{G}$ obtained by first pooling the individual posterior densities on $\boldsymbol{Y}_n \mid \boldsymbol{Y}_{\Pi_n}, \boldsymbol{X}$, then using the derived density on $\boldsymbol{X} \mid \boldsymbol{Y}^{n-1}$ to pool the individual posteriors on $\boldsymbol{Y}_{n-1} \mid \boldsymbol{Y}_{\Pi_{n-1}}, \boldsymbol{X}$, and so on. Conversely, let $\bar{f}_g$ the joint density over $\mathcal{G}$ obtained by first pooling the individual prior densities on $\boldsymbol{Y}_n \mid \boldsymbol{Y}_{\Pi_n}$ and forming the group's posterior density of $\boldsymbol{Y}_n \mid \boldsymbol{Y}_{\Pi_n}, \boldsymbol{X}$, then pooling the individual priors on $\boldsymbol{Y}_{n-1} \mid \boldsymbol{Y}_{\Pi_{n-1}}$ and forming the posterior on $\boldsymbol{Y}_{n-1} \mid \boldsymbol{Y}_{\Pi_{n-1}}, \boldsymbol{X}$ using the density of $\boldsymbol{X} \mid \boldsymbol{Y}^{n-1}$, and so on. We say that when these procedures are followed, the densities are *backward sequentially updated.* It is apparent that an operator respecting CEB is such that $\tilde{f}(\boldsymbol{y} \mid \boldsymbol{x}) = \bar{f}(\boldsymbol{y} \mid \boldsymbol{x})$.

Proposition 2.6.5. *The conditional logOp respects CEB for a decomposable CG $\mathcal{G}$ when distributions $f_i^j(\cdot)$ and $f_i^{DM}(\cdot)$, $i \in [n]$, $j \in [m]$, are backward sequentially updated.*

In Chapter 3 we discuss a variation of this result to cases where different groups of experts are responsible for disjoint subsets of the vertices of the CG. Furthermore, in Chapter 4 we develop recursions in dynamic settings which mirror the backward sequential updating of Faria and Smith [1997].

2.6.2.3   Aggregation in Complex Problems. The previous sections considered situations where experts can simply provide probabilities, possibly in a multivariate setting, for the whole domain under study. However, as we have extensively argued, in the applications we have in mind this can be only rarely the case. A first attempt to generalise such framework for expert judgement combination was proposed in Bordley [2009]. In that paper each expert delivers probabilities over a separate, but related, set of events and aggregating methods are then developed to reconcile their judgements into a unique probability distribution.

In complex applications however, experts can rarely simply deliver numbers representing their estimates and the uncertainty around these. They often run their own models and observe their outputs to then deliver the required uncertainties [French, 2011]. This should be apparent from the RODOS discussion in Chapter 1. This issue was already noticed in nuclear emergency management in Cooke and Goossens [2000]. The ENSAMBLE project [Krishnamurti et al., 2000, Mikkelsen et al., 2003] is another example of this type of combination, where the outputs of different meteorological models need to be combined into a unique weather forecast. These types of combinations introduce new challenges since correlated estimates and uncertainties may arise at both the expert and the model level.

French [2011] pointed out that the issue becomes even more delicate when models need to be chained together as in the RODOS modules' structure exemplified in Figure 1.2. Very little has been said on this issue and some early discussion on possible solutions were proposed in French et al. [1991]. In the following chapters we explicitly address these types of problems and define a framework to coherently network together models and expert judgements.

2.7   Conclusions and Contributions

This chapter has summarised the main aspects of the Bayesian approach to inference and decision making that are relevant for this thesis. We have reviewed a variety of methods to represent both the uncertainties and the preferences of individual DMs. However we have noted that current methods for groups of DMs and for combining the

judgements of several experts in complex domains are limited. In the following chapters we extend Bayesian methodology to allow for a coherent aggregation of experts' beliefs and preferences in complex domains. We are able to show that many of the graphical models we have introduced provide an intuitive basis to do this.

Note that some of the concepts introduced in this chapter, namely directional utility diagrams, MIDs and their evaluation are original contributions to the literature. We discuss additional features of these models in the following chapters.

2.7.1 Contributions

Having introduced in details the normative framework of Bayesian decision analysis and inference, we can now discuss more explicitly the contributions of this thesis to the literature. To increase readability, we discuss the extensions we propose in the same order as in the list on pages 13-14.

- in this chapter we have shown how probabilistic inference can be carried in a distributed fashion for a variety of models when certain independences are assumed for the model parameter vector (specified in Definition 2.3.9 for BNs, in Definition 2.3.21 for MNs, on page 39 for CEGs and in Proposition 2.3.39 for MDMs). In Section 3.6 we introduce weaker conditions for these models, enabling distributed inferences when different panels of experts deliver the required probabilities for disjoint subsets of the vertex set of the underlying graph;

- the combination rules for expert judgements of Faria and Smith [1997] briefly reviewed in Proposition 2.6.5 work when each expert deliver probabilities for every variable of an underlying CG. We extend this result in Section 3.6 to cases where experts deliver judgements only for subsets of the graph's vertex set;

- in Section 2.3.4 we introduced the Perlean notion of causality and discussed Perlean intervention in Definition 2.3.49. In Section 3.5 below we introduce a new notion of causality extending the one of Pearl in two ways: first, our definition does not require the existence of an idle system, but instead assumes the existence of a natural comparator - a decision rule which is standard in the domain examined; second, it allows for more general interventions than the 'do' standard operator;

- Definition 2.4.23 introduced bidirectional utility diagrams, a flexible graphical representation of partial utility independence. In Chapter 4 (and specifically Definition 4.1.8 and Proposition 4.1.10) we introduce a particular subclass of these models and deduce the utility factorization associated to this subclass. Such a

class, together with a few other assumptions often made in practice (see Section 4.1), we enables the distributed computation of expected utility $\bar{u}(\boldsymbol{d})$ of any policy $\boldsymbol{d}$ as discussed in Theorem 4.1.13;

- the computation of the expected utilities for utilities in the class mentioned in the previous point can be carried using a backward induction procedure which mirrors the routine of Faria and Smith [1997] to pool probability distributions (explicated on page 72); in Chapter 4 we introduce new backward inductions that instead of pooling probability distributions, compute optimal expected utility in multi-expert domains;

- in Section 2.1.2 we introduced moments and discusses their role in summarizing the features of a distribution. Cowell et al. [1999] and Nilsson [2001] introduced closed form recursions to compute moments of additive factorizations, as the ones in Proposition 2.4.16, over a BN model. In Chapter 5 we generalize such recursions to compute moments for multilinear factorizations introduced in Proposition 2.4.11;

- standard conditional independence, as discussed in Section 2.3.1., often implies separations between certain functions of the model's parameter that are not strictly required for the computation of the moments mentioned in the point above. In Section 5.3 we introduce new separation conditions, often implied by conditional independence, allowing for the computations of the above moments;

- we define new symbolic inferential techniques in CEG and ID models, extending current methods for discrete BNs. For these models, each entry of a probability table, θ_i say, is considered as an indeterminate of a polynomial (or equally as a variable in a computer algebra system) and inference is carried with no full numerical specification of the model's probabilities. For CEGs, we extend current methods from symmetric to asymmetric probabilistic models. For IDs, we introduce a symbolic definition of utility values and introduce the polynomial form of the expected utilities associated to these models.

Chapter 3

The Integrating Decision Support System

The probabilistic decision support techniques for single users we reviewed in Chapter 2 have many advantages: these are coherent from a Bayesian viewpoint and can embed expert judgement when necessary. However the types of demands of current applications, such as the nuclear one we reviewed in Section 1.2, require new methods to coherently combine the outputs of networked experts systems in order to provide an integrated study of the whole domain. Therefore, as argued in Chapter 1, what is needed is an integrating DSS, taking very selected probabilistic outputs from each contributing module and pasting these expert judgements together to provide a comprehensive support.

We report in this chapter the results of Smith et al. [2015] where we developed a sound statistical methodology to underpin such an integrating system. This provides a framework to faithfully encode all usable and informed expert judgements and data leading to the production of numerical scores ranking the available policies. We envisage that either any of the models we reviewed in Section 2.3 or one of the many large scale hierarchical Bayesian models of temporal spatial processes now in place [e.g. Best et al., 2005, Jewell et al., 2009, McKinley et al., 2014, Wikle et al., 2001] informs one of the components of the IDSS. Each of these modules, drawing on experimental and observational evidence supplemented by expert judgement, either directly informs arguments of a decision centre's utility function or provides the necessary input to subsequent components in the process. Different panels of experts in the domain defined by each component oversee and are responsible for the probabilistic outputs of the particular component capturing part of an underlying process. The IDSS is then designed to draw these probabilistic judgements into a single coherent picture to inform policy making. Such a coherent

IDSS would thus embody all the qualitative common knowledge and fully defensible domain knowledge of the different panels, and provide baseline judgements around which discussions and modifications could be proposed.

We describe here how and when it is possible to build such an IDSS and develop statistical methodologies to appropriately guide this knowledge integration. We first introduce a new set of axioms embedding the agreement the group of experts needs to find and, generalising results from Goldstein and O'Hagan [1996], we show that these can guarantee the existence of a coherent IDSS. We then define new group Bayesian updating routines for both observational and experimental data for cases where the IDSS is underpinned by a new notion of statistical causality. Lastly, we generalise both current independence conditions and inferential methods for single DMs to groups in a variety of graphical model classes.

The chapter is organised as follows. In Section 3.1 we briefly discuss the features of IDSSs and the possible types of support such systems can provide. Section 3.2 builds the IDSS methodology and discusses some technical properties that IDSSs need to entertain. In Section 3.3 we present some very simple illustrative examples to show the difficulties and the inconsistencies that IDSSs may exhibit. Section 3.4 presents a number of results that ensure a coherent and faithful IDSS both a priori and after the introduction of observational data. In Section 3.5 we then introduce a definition of causality tailored to IDSSs and show that experimental evidence can be accommodated in these systems whilst retaining coherence. In Section 3.6 we demonstrate that most of the models we reviewed in Sections 2.3.2 and 2.3.3 can be used as integrating overarching structure to aggregate the diverse expert systems' outputs. We conclude with a discussion.

3.1   Features of Integrating Systems

The description of the RODOS system in Section 1.2 highlighted the need in current applications for systems capable of drawing different components of the problem together. The IDSS methodology aims at supporting decision centres using expert judgement coming from different panels with different knowledge. We call the *collective* the totality of the experts in the panels, potential users and relevant stakeholders. The IDSS embeds a particular type of 'group' decision analysis where everyone in the collective accepts to delegate a particular aspect of the problem to the most informed experts only. The features of this type of decision analysis are summarised in Figure 3.1. The collective agrees on the overarching, meaning across panels, probabilistic, preferential and decision structure of the problem the IDSS addresses, as specified by the top three boxes
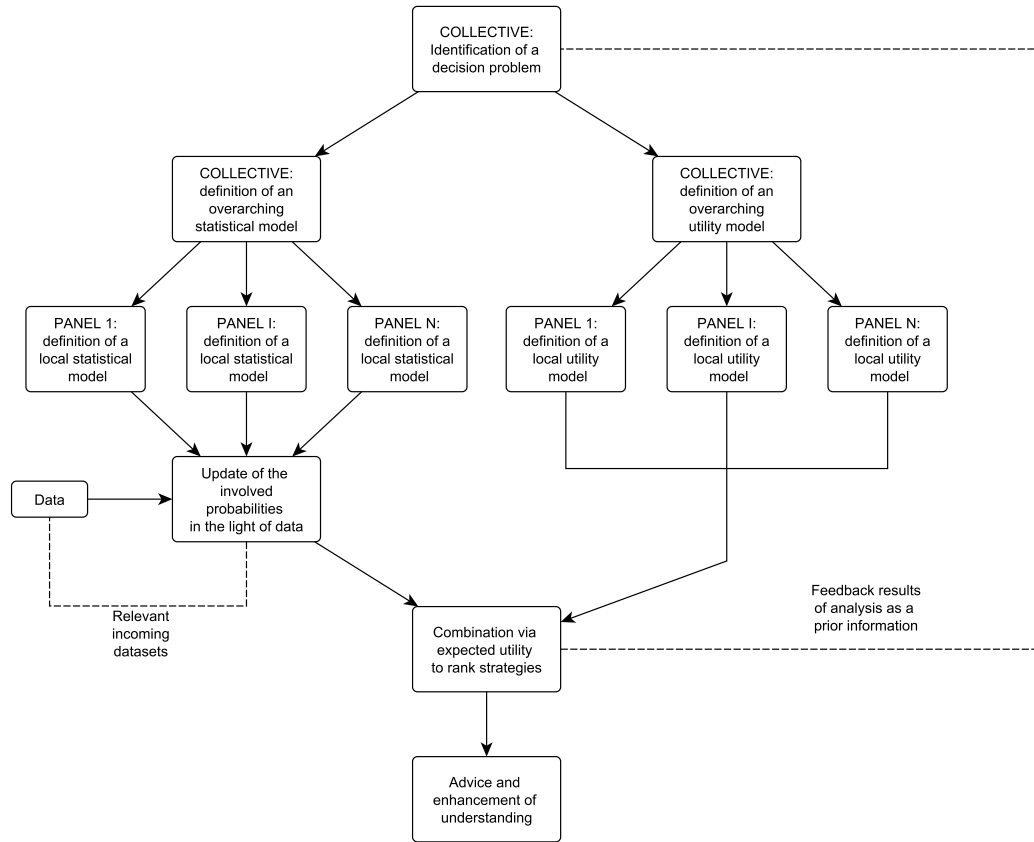
FIGURE 3.1: Structure of a Bayesian decision analysis for a group of distributed experts from Leonelli and Smith [2015], generalising French [1997].

of the diagram in Figure 3.1. By overarching statistical model, we mean that the collective identifies the main variables of the problem and the input/output relationships between these, described, for instance, by a set of conditional independence statements or by a graphical model as the network in Figure 1.2. On the other hand, a preferential overarching model consists of the identification of the attributes of the domain under study together with a set of preferential independences represented, for instance, by a specific utility factorization. In both cases no actual probability and utility elicitations are performed. Then each panel builds its own probabilistic and preferential models, only concerning the domain under its responsibility. These are then integrated into a unique entity to rank policies according to their expected utilities to provide decision support, as indicated by the bottom boxes in Figure 3.1.

One of the challenges to this process of integration is that no one person or group owns all the relevant probabilistic expert judgements. A first and essential requirement for an IDSS is that its suggestions need to be defensible to the challenges a decision centre might face to the validity of any analysis it might produce. Inputs and outputs of the subsystems and specifically the associated expressions and balancing of uncertainty

need not to be self contradictory with each other. Within a Bayesian paradigm this is a **coherence** requirement over the variables determining the overall expected utility scores, corresponding to the formal separability condition of Mahoney and Laskey [1996] discussed in Chapter 1. In a nutshell, this entails that it is possible to define a unique probability distribution over the complete system given the beliefs of the panels. For example if probability distributions are assigned by two different panels to the amount of contamination to the ground and in water respectively, which have within them an assessment of the air dispersal of contamination, then this assessment has to be the same for the two distributions. If this were not the case, a unique distribution over the IDSS could not be computed and the credibility of the composite would be clearly compromised.

A second requirement for a practical IDSS is that the panels deliver *sufficiently rich* sets of information, so that they together provide enough information to compute the expected utility function. This is in general impossible unless some assumptions are made by the collective. Note that some parts of the process under examination may be very well studied and modelled, but the knowledge about others is often patchy. So for instance, dispersal models are now well established and able to compute a variety of summaries of the spread of contamination, whilst the estimates of the political effects of the accident on the stability of the affected regions often come only in the form of expert judgement.

A third important property increasing the defensibility of the outputs of an IDSS is that this incorporates the *fullest* and *most reliable* possible *evidence* within its probability specifications. This in turn requires that expert judgements are provided only by the most informed panels. However, IDSSs that are coherent a priori might not be so any more after the introduction of data. For this reason, filters checking the data that is admitted in the IDSS to perform Bayesian updating need to be implemented. If these filters are in place, then panels can produce revised estimates of the required quantities for the decision analysis, improving the overall suggestions the system is producing and consequently supporting more focused decision making.

It is also often necessary for the IDSS to be used in real time, during the unfolding of a certain crisis. Modules are now designed so that they are fast. However this is not sufficient to guarantee that the composite as a whole is able to run quickly and still retain its coherence. To achieve this, a necessary property is that the system is **distributed** (or *modular*). By this we mean that the composite system calculates all functions needed to evaluate the expected utility associated with its available decisions from outputs delivered by individual panels, where these are allowed to be functions

of other panels' outputs. There are several advantages that derive from structuring a problems so that the ensuing support is distributed:

1. first, because the responsibility for each aspect of the analysis can be devolved to *appropriate* panels of experts, these are then more likely to deliver better calibrated judgements. This in some ways guarantees the semantic separability of Mahoney and Laskey [1996]. The whole system might therefore be expected to be more robust to the misspecification of beliefs [Cooke, 1991]. On the other hand, if the system needs to be changed in the light of unexpected developments or unplanned consequences, under suitable conditions, the management of these new developments need only be addressed *autonomously* by the relevant panels. These simply adapt their individual forecasts and inputs in the light of the new scenario they face. These adjustments can then be folded back into the system to provide a revised output of the relevant modules for other panels to use for their inputs;

2. second, the output of a distributed IDSS can produce answers to queries by decision centres about the premises on which it is based and the calculations of its outputs, by directing the query to the relevant panels. A route along which the relevant panels can be queried as to the reasons of their contribution to the expected utility scores is described in Figure 3.2, which shows the different levels of support that can be implemented into an IDSS. When queried, generic DSSs are usually able to present justifications for the overall expected utility scores and the ranking of the available policies (third box of the second column from the left of Figure 3.2). However IDSSs are able to provide an additional level of support. The system's distributivity permits the IDSS to justify its suggestions in terms of the modules outputs, since expected utilities are functions of these (see the bottom box of the second column from the left of Figure 3.2). If these systems did not exhibit this distributivity property, then such devolution may not be possible and so any support would be much less transparent. This feature increases both the comprehensibility and the feasibility of testing that, as argued by Mahoney and Laskey [1996], current DSSs need to address;

3. third, distributivity ensures that different decision centres can be given the option of choosing different modules to model the various components of the problem. For example, in nuclear emergency management, different countries often prefer to predict the spread of the contamination using their national agencies' diffusion models;

4. lastly, although many of the modules are probabilistic, distributivity allows us not to construct a single composite probabilistic model: such a universal system would
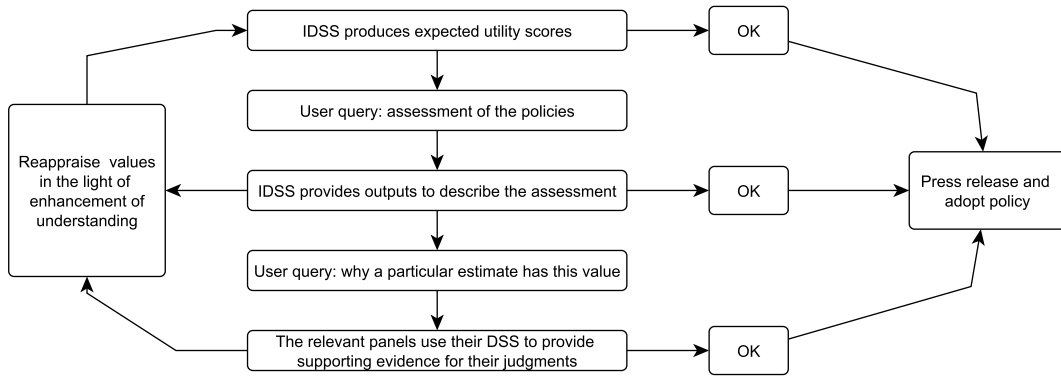
FIGURE 3.2: Description of the possible use of an integrating decision support system
for a decision analysis.

be infeasibly large and no one would own the joint distributions of the unwieldy
number of random variables. Even if it were possible to build such a system, the
component modules are usually being constantly revised both in their structure
and inputs. Were such an overarching probabilistic system possible to build, it
would therefore be obsolete before it was completed.

3.2 Construction of Integrating Systems

3.2.1 Some Technical Structure

A domain is defined by a large number of random variables $\boldsymbol{Y} = (Y_i)_{i \in [n]}^{\mathrm{T}}$. As often in
practice, the problem is heterogeneous and therefore assume different components of the
problem, denoted as $\boldsymbol{Y}_i$, are evaluated and overseen by m different panels $\{G_i : i \in [m]\}$
of domain experts, $[m] = \{1, \ldots, m\}$. The implicit, albeit virtual, owner of these beliefs
is henceforth referred to as the SB. Let $\boldsymbol{\mathcal{D}}$ be a decision space including the available
policies $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$. The SB needs to process the necessary probabilistic features delivered by
the different panels to calculate various statistics of a potential decision centre's reward
vector $\boldsymbol{R}$, some function of both $\boldsymbol{Y}$ and $\boldsymbol{d}$.

Panel G_i donates to the SB various summaries of the distribution of the subvector $\boldsymbol{Y}_i \mid \boldsymbol{d}$
under its jurisdiction, conditional on certain measurable functions $\{L_i(\boldsymbol{Y}_i) : L_i \in \mathcal{L}_i\}$,
where $\mathcal{L}_i$ could be empty, $i \in [m]$. For instance, in a generic graphical model the set $\mathcal{L}_i$
might consist of the possible configurations of values of the parents of $\boldsymbol{Y}_i$. Assume that
panel G_i, $i \in [m]$, is able to make available the following:

1. a collection of summary statements about the measurements under its jurisdiction

$$\Psi_i^{\boldsymbol{y}|\boldsymbol{\theta}} = \{\Psi_i^{\boldsymbol{y}|\boldsymbol{\theta}}(\boldsymbol{\theta}_i, L_i, \boldsymbol{d}) : \boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i, \ L_i \in \mathcal{L}_i, \ \boldsymbol{d} \in \mathcal{D}\}.$$

These inform the IDSS about the likelihoods $f_i(\boldsymbol{y}_i \mid \boldsymbol{\theta}_i, L_i, \boldsymbol{d})$, $L_i \in \mathcal{L}_i$, $\boldsymbol{d} \in \mathcal{D}$, where $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i$ parametrise the (possibly conditional) density of $\boldsymbol{Y}_i \mid \boldsymbol{\theta}, \boldsymbol{d}$. For example, if $\boldsymbol{Y}$ were discrete and finite, then each panel might be asked to provide certain contingency tables over $\boldsymbol{Y}_i \mid \boldsymbol{\theta}, \boldsymbol{d}$, conditional on each $L_i \in \mathcal{L}_i$. In this case $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i$ would be the probabilities within all these tables. An important note to make here is that there is often a typically much longer vector of parameters $\boldsymbol{\phi}_i \in \boldsymbol{\Phi}_i$ over which G_i has beliefs but which are not directly relevant to the distribution of the predictions the IDSS needs to deliver;

2. a set of summaries about its prior beliefs

$$\Psi_i^{\boldsymbol{\theta}} = \left\{\Psi_i^{\boldsymbol{\theta}}(L_i, \boldsymbol{d}) : L_i \in \mathcal{L}_i, \ \boldsymbol{d} \in \mathcal{D}\right\},$$

from the panel prior densities $\pi_i(\boldsymbol{\theta}_i \mid L_i, \ \boldsymbol{d})$, $\boldsymbol{d} \in \mathcal{D}$, $L_i \in \mathcal{L}_i$. In the above example this would be a joint probability distribution over the entries of the relevant contingency tables. Note that often these panel beliefs are calculated by marginalising out $\boldsymbol{\phi}_i \in \boldsymbol{\Phi}_i$;

3. from these two sets of quantities, the calculated collection of summaries

$$\Psi_i^{\boldsymbol{y}} = \left\{\Psi_i^{\boldsymbol{y}}(L_i, \boldsymbol{d}) : L_i \in \mathcal{L}_i, \ \boldsymbol{d} \in \mathcal{D}\right\},$$

about the marginal distribution of $\boldsymbol{Y}_i \mid \boldsymbol{d}$ conditional on each event L_i. Here $\Psi_i^{\boldsymbol{y}}(L_i, \boldsymbol{d})$ is calculated from $\Psi_i^{\boldsymbol{y}|\boldsymbol{\theta}}(\boldsymbol{\theta}_i, L_i, \boldsymbol{d})$ by G_i marginalising over $\boldsymbol{\theta}_i$ using $\Psi_i^{\boldsymbol{\theta}}(L_i, \boldsymbol{d})$ for each $L_i \in \mathcal{L}_i$, $\boldsymbol{d} \in \mathcal{D}$. In the example above, these might be the corresponding expected conditional tables where expectations are taken across the probability vectors $\boldsymbol{\theta}_i$. These are the quantities the IDSS uses to calculate its expected utility scores associated with its available decisions and to help a decision centre to determine its most efficacious policy.

3.2.2 Common Knowledge Axioms

Suppose that after a series of decision conferences (see Section 2.6.1) held jointly across the panels, stakeholders and potential users, all have agreed the types of decisions the IDSS supports to a sufficient level of specificity to provide an agreed qualitative framework across all interested parties around which a quantitative structure can be built. To this purpose we make three assumptions.

Axiom 3.2.1 (policy consensus)**.** *The collective agrees the class of decision rules $\boldsymbol{d} \in \mathcal{D}$ examined by the IDSS.*

This class of feasible policies considered depends not only on what is logical, such as when various pieces of information are likely to become available, but also what might be allowable, either legally or for other reasons.

Axiom 3.2.2 (utility consensus)**.** *The collective agrees on the class $\mathcal{U}$ of utility functions supported by the IDSS.*

In the complex multivariate settings we address here, the utility function $u(\boldsymbol{r}, \boldsymbol{d})$ needs to entertain certain types of preferential independence in order to allow for a distributed analysis. In this chapter we assume that some additive or multilinear factorisation is believed to hold. In Chapter 4 we introduce a large class of partial utility independence models, of which additive and multilinear factorisations can be seen as a special case, that can allow for distributed analyses. Again the choice of $\mathcal{U}$ is often resolved using decision conferencing across the members of the collective.

Axiom 3.2.3 (structural consensus)**.** *The collective agrees the variables $\boldsymbol{Y}$ defining the process - where for each $\boldsymbol{d} \in \mathcal{D}$, each $u \in \mathcal{U}$ is a function of $\boldsymbol{Y}$ - together with a set of qualitative statements about the dependence between various functions of $\boldsymbol{Y}$, $\boldsymbol{\theta}$ and $\boldsymbol{d}$. Call this set of assumptions the **structural consensus set** and denote this by $\mathcal{J}$.*

This last consensus might be expressible through agreement that a particular graphical or conditional independence structure across not only the distribution of $\boldsymbol{Y} \mid \boldsymbol{\theta}, \boldsymbol{d}$, but also over the one of $\boldsymbol{\theta} \mid \boldsymbol{d}$ is valid, $\boldsymbol{d} \in \mathcal{D}$. For instance, the structural consensus might include the assumption of local and global independence of the parameter vector in a BN model (see Section 2.3.2.2). Other information that might be included in $\mathcal{J}$ could be a consensus about certain structural zeros or known logical constraints (such as probabilities needing to add to one and be non-negative).

Definition 3.2.4. *Call the set of assumptions forming the union of the utility, policy and structural consensus $(\mathcal{U}, \mathcal{D}, \mathcal{J})$ the **Common Knowledge-class (CK-class))**.*

Technically, we can think of the CK-class as the *qualitative* beliefs that are shared as common knowledge by all panel members and potential users. The CK-class represents the foundation around which all inference within the IDSS takes place. Note that this depends not only on the domain and needs of users of the system, but also on the constitution and knowledge bases of the panels.

We now provide a list of the types of quantitative information needed to populate the CK-class and to calculate the expected utility $\bar{u}(\boldsymbol{d})$ of the available decisions $\boldsymbol{d} \in \mathcal{D}$.

To achieve this, the SB needs the decision centre's choice of $u(\boldsymbol{r}, \boldsymbol{d}) \in \mathcal{U}$ together with enough probabilistic information to calculate the expectations of these utilities. At worst this might need to be the full distribution of $\boldsymbol{R}$. Alternatively and more commonly for typical choices of $\mathcal{U}$, all that might be needed is the distribution of the margins on certain specific functions of $\boldsymbol{R}$ or simply some summaries.

Axiom 3.2.5 (quantitative delegation consensus)**.** *The collective agrees to take on the sample summaries* $\Psi_i^{\boldsymbol{y}|\boldsymbol{\theta}}$, *the panel prior beliefs* $\Psi_i^{\boldsymbol{\theta}}$ *and the panel marginal inputs* $\Psi_i^{\boldsymbol{y}}$ *delivered by G_i as its own, $i \in [m]$.*

This axiom essentially demands that everyone agrees that it is appropriate to defer their judgements to the panel which is most informed about each domain vector. A sufficient set of qualitative conditions justifying its use is given in Section 3.4.1.

3.2.3 Properties of Integrating Systems

The above axioms specify the structure of the IDSS, but these do not in general guarantee that the output of the system provides coherent decision support. In this section we introduce concepts that characterise what a 'good' IDSS is.

Definition 3.2.6. *An IDSS is **adequate** for a CK-class if the SB can unambiguously calculate the expected utility scores, for any decision $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$ and any utility function $u \in \mathcal{U}$, from the panel marginal inputs $\Psi_i^{\boldsymbol{y}}$ delivered by G_i, $i \in [m]$.*

One particular useful refinement of adequacy is the following.

Definition 3.2.7. *An IDSS is **universally adequate** for a CK-class if the SB can unambiguously calculate the distribution of $\boldsymbol{R}$ from the panel marginal inputs $\Psi_i^{\boldsymbol{y}}$ delivered by G_i, $i \in [m]$.*

An adequate IDSS is able to derive a unique score for each option on the basis of the individual panels' inputs. An IDSS clearly cannot be fully functional unless it has this property. For a universally adequate IDSS this is true regardless of the form of $\mathcal{U}$.

To be defensible the IDSS needs another property.

Definition 3.2.8. *An IDSS is **sound** for a CK-class if it is adequate and the SB can coherently admit all the assessments $\Psi_i^{\boldsymbol{y}|\boldsymbol{\theta}}, \Psi_i^{\boldsymbol{\theta}}$ and $\Psi_i^{\boldsymbol{y}}$, $i \in [m]$, as her own, the SB's underlying belief model being shared with those of the relevant panels.*

A sound IDSS does not necessarily need to embody the *genuine* beliefs held by the panel members and potential users based on the totality of their own personal evidence. This would be too much to ask since much of this evidence might be from poorly designed experiments, formally unjustifiable to the public or simply anecdotal. However the sound IDSS does present a defensible and conservative position all panellist should be happy to communicate, and provide a benchmark for further discussion. Most importantly soundness guarantees that the beliefs are formally separable.

A property we always assume to be part of a CK-class is that panels are variationally independent, i.e $\boldsymbol{\Theta} = \times_{i \in [m]} \boldsymbol{\Theta}_i$ (see Definition 2.3.11). If this were not the case, then beliefs of one panel would be necessarily shared with the ones of other panels. In such case both soundness and distributivity could not hold for IDSSs.

3.3 Illustrative Examples

In this section we present a series of simple and intuitive examples illustrating the IDSS machinery and the dangers associated to non coherent belief specifications.

3.3.1 Strong Adequacy and Soundness a Priori.

Let $m = 2$, $\boldsymbol{R} = \boldsymbol{Y} = (Y_1, Y_2)^{\mathrm{T}}$, where Y_i is binary, $i \in [2]$, $\boldsymbol{\theta}$ parametrise the density of $(Y_1, Y_2) \mid \boldsymbol{\theta}, \boldsymbol{d}$, and suppose a generic decision space $\boldsymbol{\mathcal{D}}$. Here the random variable Y_1 is an indicator of whether or not the contamination is dispersed in the area and Y_2 is an indicator of whether or not the population is affected by the accident. The sample summary delivered by G_1 is $\theta_1 = \mathbb{P}(Y_1 = 1 \mid \theta_1, \boldsymbol{d})$, whilst G_2 delivers $\{\theta_{20}, \theta_{21}\}$, where $\theta_{20} = \mathbb{P}(Y_2 = 1 \mid Y_1 = 0, \theta_{20}, \boldsymbol{d})$ and $\theta_{21} = \mathbb{P}(Y_2 = 1 \mid Y_1 = 1, \theta_{21}, \boldsymbol{d})$. Write $\boldsymbol{\theta}_2 = (\theta_{20}, \theta_{21})^{\mathrm{T}}$. Note that in this case $\mathcal{L}_1$ is empty, whilst $\mathcal{L}_2$ consists of the possible outcomes of Y_1.

If a CK-class includes $\mathcal{U}$, an arbitrary utility on $\boldsymbol{R}$, then, since the above probabilities fully define the model, for strong adequacy the SB needs to be able to calculate the expected joint probability table of $\boldsymbol{Y} \mid \boldsymbol{d}$. Specifically this corresponds to the expected values $\bar{\boldsymbol{\mu}} = (\bar{\mu}_{00}, \bar{\mu}_{01}, \bar{\mu}_{10}, \bar{\mu}_{11})^{\mathrm{T}}$ of $\bar{\boldsymbol{\theta}} = (\bar{\theta}_{00}, \bar{\theta}_{01}, \bar{\theta}_{10}, \bar{\theta}_{11})^{\mathrm{T}}$, where by definition

$$\bar{\theta}_{00} = (1 - \theta_1)(1 - \theta_{20}), \quad \bar{\theta}_{01} = (1 - \theta_1)\theta_{20}, \quad \bar{\theta}_{10} = \theta_1(1 - \theta_{21}), \quad \bar{\theta}_{11} = \theta_1\theta_{21}.$$

Suppose that the qualitative statement $\theta_1 \perp\!\!\!\perp \boldsymbol{\theta}_2 \mid \boldsymbol{d}$ is in the CK-class for every $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$. Letting $\mu_1 = \mathbb{E}(\theta_1 \mid \boldsymbol{d})$ and $\boldsymbol{\mu}_2 = (\mu_{20}, \mu_{21})^{\mathrm{T}} = (\mathbb{E}(\theta_{20} \mid \boldsymbol{d}), \mathbb{E}(\theta_{21} \mid \boldsymbol{d}))^{\mathrm{T}}$, we then have

that

$$\bar{\mu}_{00} = (1 - \mu_1)(1 - \mu_{20}), \quad \bar{\mu}_{01} = (1 - \mu_1)\mu_{20}, \quad \bar{\mu}_{10} = \mu_1(1 - \mu_{21}), \quad \bar{\mu}_{11} = \mu_1\mu_{21}.$$

Suppose G_1 and G_2 make available their expectations μ_1 and $\boldsymbol{\mu}_2$, respectively. Then, because of the independence statements between the parameter vectors in the structural consensus, the IDSS is a strongly adequate system with the delivered assessments, providing the SB with all the information needed to calculate the expected utility functions using the formulae in (3.3.1). It is also sound since these inputs are consistent with any probability model over $(\boldsymbol{Y}, \theta_1, \boldsymbol{\theta}_2) \mid \boldsymbol{d}$ with these expectations and the independence between the parameter vectors.

It is not a trivial condition that SB's beliefs required for either adequacy or strong adequacy are a function of the panels beliefs only. For example the full joint distribution of $\boldsymbol{Y} \mid \boldsymbol{\theta}, \boldsymbol{d}$ is not fully recoverable from the marginal densities over $\theta_1 \mid \boldsymbol{d}$ and $\boldsymbol{\theta}_2 \mid \boldsymbol{d}$ two different panels might simply provide, since the SB cannot derive the conditional covariance between θ_1 and $\boldsymbol{\theta}_2$ given $\boldsymbol{d}$ which was needed to calculate the conditional covariance between Y_1 and Y_2 given $\boldsymbol{d}$.

3.3.2 Adequacy and the Risk of non Soundness.

Assume a CK-class gives $\boldsymbol{Y}$ the same meaning as in the previous section. However add to the CK-class the additional structural assumption that $Y_2 \perp\!\!\!\perp Y_1 \mid \theta_1, \boldsymbol{\theta}_2, \boldsymbol{d}$ whatever decision $\boldsymbol{d} \in \mathcal{D}$ is made. Thus, once the probabilities of these events are known, it is generally accepted that learning that contamination has been introduced in the environment does not affect the judgements about the health effects on the population. Note that in this case $\mathcal{L}_i$ is empty, $i \in [2]$. Suppose G_1 delivers the beta distributions $\mathrm{Be}(p_1, q_1)$ for $\theta_1 = \mathbb{P}(Y_1 = 1 \mid \theta_1, \boldsymbol{d})$ and G_2 has the beta distributions $\mathrm{Be}(p_2, q_2)$ for $\theta_2 = \mathbb{P}(Y_2 = 1 \mid \theta_2, \boldsymbol{d})$. Note that because of the structural assumption above, our notation is now such that $\theta_2 = \theta_{20} = \theta_{21}$. Consider two possible CK-classes where a decision centre is known to draw its utilities $u_i \in \mathcal{U}_i$, $i \in [2]$, from one of the families below

$$u_1(y_1, y_2, \boldsymbol{d}) = a_1 + b_{11}y_1 + b_{12}y_2,$$
$$u_2(y_1, y_2, \boldsymbol{d}) = a_2 + b_2 y_1 y_2,$$

where $a_1, a_2 \in \mathbb{R}$ and $b_{11}, b_{12}, b_2 \in \mathbb{R}_{>0}$. If $\mathcal{U}_1$ is in the CK-class then the SB needs only G_i to supply its prior mean μ_i of $\theta_i \mid \boldsymbol{d}$, $i \in [2]$, where $\mu_i = p_i(p_i + q_i)^{-1}$, in order to calculate the expected utility associated to $\boldsymbol{d}$. So there is some redundancy in the

delivery of G_1 and G_2: panels only need to specify their prior means for every $\boldsymbol{d} \in \mathcal{D}$ and not their whole marginal distributions over $\theta_1 \mid \boldsymbol{d}$ and $\theta_2 \mid \boldsymbol{d}$.

However, if $\mathcal{U}_2$ is in the CK-class, then the SB needs to be able to calculate

$$\mathbb{E}(Y_1 Y_2 \mid \theta_1, \theta_2, \boldsymbol{d}) = \mathbb{E}\left(\theta_1 \theta_2 \mid \boldsymbol{d}\right),$$

and the panels prior means are no longer necessarily adequate. For this to be so the CK-class then needs the additional assumption $\theta_1 \perp\!\!\!\perp \theta_2 \mid \boldsymbol{d}$. In this case

$$\mathbb{E}(Y_1 Y_2 \mid \theta_1, \theta_2, \boldsymbol{d}) = \mathbb{E}(\theta_1 \theta_2 \mid \boldsymbol{d}) = \mu_1 \mu_2,$$

and the IDSS retrieves adequacy. However, this additional common knowledge statement needs to be credible. For instance, suppose that $p_1 + q_1 = p_2 + q_2 = \sigma$, so that $\Theta_1 \times \Theta_2$ is parametrised by (μ_1, μ_2, σ). Then, it is easily checked that the collective can specify its joint beliefs over the probabilities on the 4 events of the form (y_1, y_2), $y_1, y_2 = [1]^0 = [1] \cup \{0\}$, with a Dirichlet distribution, $\mathrm{Di}(a_{00}, a_{10}, a_{01}, a_{11})$, where

$$p_1 = a_{10} + a_{11}, \quad q_1 = a_{00} + a_{01}, \quad p_2 = a_{01} + a_{11}, \quad q_2 = a_{00} + a_{10},$$

since, from well known properties of the Dirichlet distribution [see e.g. Geiger and Heckerman, 1997], this is consistent with the given margins. However this collective prior is *not* consistent with the independence assumption $\theta_1 \perp\!\!\!\perp \theta_2 \mid \boldsymbol{d}$ given above. From the properties of the Dirichlet (see Appendix C.1.2), the SB's mean of $\theta_1 \theta_2 \mid \boldsymbol{d}$ is given by $a_{11} \sigma^{-1}$ where $\sigma = a_{00} + a_{10} + a_{10} + a_{11}$, which is not equal to $\mu_1 \mu_2$ unless $\rho = \sigma^{-2}\left(a_{11} a_{00} - a_{10} a_{11}\right) = 0$, since ρ is the unique solution to

$$\mu_1 \mu_2 = a_{11} \sigma^{-1} \iff \frac{(a_{10} + a_{11})(a_{01} + a_{11})}{\sigma^2} = \frac{a_{11}}{\sigma}.$$

So the SB cannot calculate $\mathbb{E}(\theta_1 \theta_2 \mid \boldsymbol{d})$ from the inputs delivered by the panels. In fact $\mathbb{E}(\theta_1 \theta_2 \mid \boldsymbol{d}) = \mu_1 \mu_2 + \rho$, which is unidentified from the margins provided by the panels.

3.3.3 Bayesian Updating

Ideally we would like the IDSS to be distributed so that panels can autonomously update their probabilistic beliefs as they receive new information. To illustrate how distributivity might be possible, suppose a random vector $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ is sampled from the same population as (Y_1, Y_2) in the model above and that, for each $\boldsymbol{d} \in \mathcal{D}$, $\theta_1 \perp\!\!\!\perp \theta_2 \mid \boldsymbol{d}$ is in the CK-class. Each panel G_i next refines its probabilistic assessments by observing

its own separate randomly sampled populations, $\boldsymbol{x}_i$, and then updates its beliefs, given each $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$, from $\pi_i(\theta_i \mid \boldsymbol{d})$ to $\pi_i(\theta_i \mid \boldsymbol{d}, \boldsymbol{x}_i)$, $i \in [2]$.

Now the two panels need to deliver only their respective posterior means $\mathbb{E}(\theta_i \mid \boldsymbol{d}, \boldsymbol{x}_i)$, $i \in [2]$, $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$. The SB can then act coherently and as if she had processed this information on her own: the IDSS is therefore sound and distributed. So in particular once all the data accommodated into the composite is of the form of autonomous random sampling, inference can be simply delegated to the appropriate panels who simply periodically and independently revise their judgements.

However note that parameter independence is critical for this distributivity property. Assume the SB uses the additive utility class $\mathcal{U}_1$, but that there are no clear reasons why $\theta_1 \perp\!\!\!\perp \theta_2 \mid \boldsymbol{d}$ should be in the CK-class, so that the prior over θ_2 needs to be a function of θ_1 for at least some $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$. Then with these beliefs, were the SB a single agent, she would for example draw on what she learns about θ_1 from $\boldsymbol{x}_1$ to update her beliefs about θ_2. So her posterior density on $\theta_2 \mid \boldsymbol{d}$

$$\pi_2(\theta_2 \mid \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{d}) = \int_0^1 \pi_2(\theta_2 \mid \theta_1, \boldsymbol{x}_2, \boldsymbol{d})\pi_1(\theta_1 \mid \boldsymbol{x}_1, \boldsymbol{d})\mathrm{d}\theta_1,$$

has associated expectation $\mathbb{E}(\theta_2 \mid \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{d}) \neq \mathbb{E}(\theta_2 \mid \boldsymbol{x}_2, \boldsymbol{d})$ in general. Therefore the SB using $\mathbb{E}(\theta_i \mid \boldsymbol{x}_i, \boldsymbol{d})$, $i \in [2]$, will *not* be acting as a single Bayesian would. So the system is no longer sound. Although when supporting evidence remains unseen the SB appears to act coherently, her analyses are indefensible if subsequently challenged.

3.3.4 Learning with Randomised Samples.

Note that even if the parameter independence $\theta_1 \perp\!\!\!\perp \theta_2 \mid \boldsymbol{d}$ is justified a priori, the assumption that data collected by the two panels and individually used to adjust their beliefs does not inform both parameters is also a critical one. Continuing the example above and assuming $u \in \mathcal{U}_2$, suppose that G_1 and G_2 both see the results of the experiment in Table 3.1, where 100 units from the population are randomly sampled, for a particular $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$. Each panel uses this experiment to update its respective marginal distributions over $\theta_i \mid \boldsymbol{d}$, $i \in [2]$.

Then, if both began with a prior symmetric about 0.5, each would believe that its posterior mean would still be equal to one half. So were $\theta_1 \perp\!\!\!\perp \theta_2 \mid \boldsymbol{d}$ in the CK-class and assuming that all evidence the individual panels use is from the marginal counts only, the IDSS would assign $\mathbb{E}(\theta_1\theta_2 \mid x_1, x_2, \boldsymbol{d}) = 0.25$. In contrast, were the SB to see this table, assuming $\theta_1 \perp\!\!\!\perp \theta_2 \mid \boldsymbol{d}$ a priori with fairly uninformative priors on the two margins, then her posterior mean of $\theta_1\theta_2 \mid \boldsymbol{d}$ would be approximately 0.05, which

Y_1/Y_2	0	1		
0	5	45	50	$n - x_1$
1	45	5	50	x_1
	50	50	100	
	$n - x_2$	x_2		

TABLE 3.1: A random sampled experiment for the example in Section 3.3.4.

is way different than the estimate above. No one may have access to the joint table but just its marginal counts. So unless a protocol is adopted by the IDSS preventing data inducing this type of ambiguity into the distributed IDSS, then this system might be grossly misleading. Interestingly if the CK-class does not include $Y_2 \perp\!\!\!\perp Y_1 \mid \theta_1, \theta_2$, then we can prove below that in the extended system there is a simple solution to this problem. This is because the given sample data appears to challenge the assumption $Y_2 \perp\!\!\!\perp Y_1 \mid \theta_1, \theta_2, \boldsymbol{d}$. When an observed likelihood destroys the prior independence of the parameters in the two posterior margins, the useful distributive property is lost because the independence between θ_1 and θ_2 given $\boldsymbol{d}$ no longer holds a posteriori.

So we have illustrated above that even in the simplest of networks, considerable care needs to be exercised before an IDSS can be expected to work reliably. In the next section we prove some conditions which ensure an IDSS is sound both a priori and a posteriori.

3.4 Conditions for a Coherent Integrating System

Having illustrated above some of the challenges faced by an IDSS, in this section we investigate sufficient conditions leading to coherence and distributivity. We are able to demonstrate that if panels are constructed wisely and that care is taken in defining what observational data is allowed to inform IDSSs, the necessary conditions are not overly restrictive.

However, an IDSS needs to have an appropriate protocol which systematically excludes or delays some pieces of information. Suppose a sequence of datasets is presented to the IDSS as time progresses. We call the information that such a protocol allows into the system up to time t the *admissible evidence* and denote it by I_+. This mirrors the information set in the definition of the DLM model class of Section 2.3.3.1. The types of information excluded or delayed by this protocol are simply those that either cause ambiguity or lack of consensus and therefore break the distributivity of the IDSS. In this sense the beliefs of the IDSS are conservative. Of course all statements in an IDSS simply provide a benchmark from which further discussion - possibly involving more

contentious sources of evidence - can ensue before acts are decided. But we argue that primarily an IDSS needs to be able to deliver outputs based on considered and agreed inputs from the expert panellist and that its judgements have a supporting narrative associated with them, which can be appraised and, if necessary, replaced during any given crisis as described in Figure 3.2.

Of course, that there exist relevant protocols for the selection of good quality evidence for decision support is often assumed even for single agent systems, but its explicit statement is frequently omitted. For instance, the Cochrane reviews are considered to be the gold standard in decision support for medical treatments [Higgins and Green, 2008]. Their purpose is to pare away information which might be ambiguous and potentially distort inferences through a highly-developed and trusted set of principles relevant to the domain. Here, we assume that panels can select suitable evidence, mirroring Cochrane in ways relevant to their domain. We also assume that such relevant analysis can be performed locally. We demonstrate below that conditions justifying this strongly relate to certain conditional independence statements leading to the separability across different panel parameters of the observed likelihood of portfolios of data.

3.4.1 Group Conditional Independences

To deduce these independences, let $I_{CK}^t \subseteq I_+^t$ denote all the admissible evidence which is common knowledge to all panel members at time t and I_{ij}^t denote the subset of I_+^t panel G_i would use at time t if acting autonomously to assess its beliefs about $\boldsymbol{\theta}_j$, $i, j \in [m]$. Therefore $\cup_{i,j\in[m]} I_{ij}^t \subseteq I_+^t$. Further let $I_*^t = \cup_{j\in[m]} I_{jj}^t$.

The issue now is to determine what might constitute good criteria for determining whether or not certain data can be admitted into the IDSS. One very convenient property to demand - which, with other assumptions, ensures that the IDSS continues to be unambiguous and distributed - is the following.

Definition 3.4.1. *The IDSS exhibits **panel independence** with respect to I_+^t at time t if $\perp\!\!\!\perp_{i\in[m]} \boldsymbol{\theta}_i \mid I_+^t, \boldsymbol{d}$, for every $\boldsymbol{d} \in \mathcal{D}$.*

It should be noted that in some contexts panel independence is a strong assumption. This is most often violated when a panel has marginalised out various indirect explanatory variables, $\boldsymbol{\phi}_i \in \Phi_i$, used to build G_i's model and some of these are dependent on the indirect variables $\boldsymbol{\phi}_j \in \Phi_j$ of another panel G_j, $i, j \in [m]$. In this case marginalisation can induce a sometimes quite strong prior dependence between $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$. The induced dependence has close links with the presence of unobserved confounders [Greenland et al., 1999] in the composite system.

Let $\boldsymbol{\theta}_{i^-} = (\boldsymbol{\theta}_j)_{j \in [m] \setminus \{i\}}^{\mathrm{T}}$. Four other conditions are convenient to impose.

Definition 3.4.2. *For any $\boldsymbol{d} \in \mathcal{D}$ and $i \in [m]$, we say that a CK-class of an IDSS is* delegatable *at time t if*

$$I_+^t \perp\!\!\!\perp \boldsymbol{\theta} \mid I_{CK}^t, I_*^t, \boldsymbol{d}, \tag{3.4.1}$$

separately informed *at time t if*

$$I_{ii}^t \perp\!\!\!\perp \boldsymbol{\theta}_{i^-} \mid I_{CK}^t, \boldsymbol{\theta}_i, \boldsymbol{d}, \tag{3.4.2}$$

cutting *at time t if*

$$I_*^t \perp\!\!\!\perp \boldsymbol{\theta}_i \mid I_{CK}^t, I_{ii}^t, \boldsymbol{\theta}_{i^-}, \boldsymbol{d}, \tag{3.4.3}$$

and commonly separated *at time t if*

$$\perp\!\!\!\perp_{i \in [m]} \boldsymbol{\theta}_i \mid I_{CK}^t, \boldsymbol{d}. \tag{3.4.4}$$

If an IDSS is delegatable at time t, then everyone agrees that the totality of admissible evidence I_+^t fed into the IDSS is the union of evidence shared by all panels I_{CK}^t plus the individual evidence I_*^t each panel has about its own domain at that time. If the system is not delegatable, then clearly no consensus across the panels could be achieved. When a system is separately informed, evidence G_i might collect individually is not informative about the parameters owned by other panellists once the evidence shared between panels has been fed in. When a system is cutting, once I_{CK}^t and I_{ii}^t are known, no one in the collective believes that any panel has available information that G_i might also want to use to adjust its beliefs about $\boldsymbol{\theta}_i$. So if for instance another panel, G_j say, might have needed to marginalise out a parameter in $\boldsymbol{\theta}_i$ to accommodate a piece of evidence because its sample distribution depended on this component, then the IDSS would not be cutting: this evidence would have told the SB not only about $\boldsymbol{\theta}_j$ but also $\boldsymbol{\theta}_i$. When parameters are commonly separated all the shared information separates the parameters in the system. For example at time 0 if our overarching structure were a BN, then this condition would be satisfied if we had prior global independence of the parameter vector (see Definition 2.3.9). It would then continue to be satisfied if the protocol ensured that the global independence of these parameters were preserved a posteriori as in Proposition 2.3.12.

A CK-class including the above four properties is guaranteed to be sound and distributed, as specified by the following theorem, analogous to Goldstein and O'Hagan [1996] about the use of linear Bayes in single agent systems.

Theorem 3.4.3. *Suppose an IDSS for a CK-class $(\mathcal{U}, \mathcal{D}, \mathcal{J})$ is adequate, where $\mathcal{U}$ and $\mathcal{D}$ are arbitrary and $\mathcal{J}$ includes the consensus that the IDSS is delegatable, separately informed, cutting and commonly separated at time t. Then the IDSS is also sound and distributed at time t.*

The proof of this result is presented in Appendix A.2.1.

So a structural consensus $\mathcal{J}$ including an admissibility protocol respecting the four conditions in Definition 3.4.2 gives rise to a sound system where the SB (and all panels) assumes panel independence always holds (see the proof of Theorem 3.4.3 in Section A.2.1 for more details). Moreover the system remains distributed. Although these conditions could not by any stretch be called 'regularity conditions', they are nevertheless satisfied by a very diverse collection of models, as we illustrate in Section 3.6. In particular, being irrelevance statements, these are qualitative in nature rather than quantitative and so realistic candidates for being included in a CK-class. Note that this theorem applies to universally adequate systems, weaker conditions are needed for simply adequate systems: see Chapter 5 for more details on this issue.

3.4.2 Likelihood Separation

We now focus on the introduction of observational data in the IDSS in order to perform Bayesian updating, and consequently more focused decision making. We are able to deduce what conditions can ensure a sound and distributed IDSS to be so also a posteriori. For this purpose, assume that the only evidence presented to the IDSS is in the form of data sets $\boldsymbol{x}^t = \{\boldsymbol{x}_\tau : \tau \leq t\}$ which periodically become available to panels to populate I_+^t. Let $l(\boldsymbol{\theta} \mid \boldsymbol{x}^t)$, $t \geq 0$, denote a likelihood over the parameter $\boldsymbol{\theta}$ of the distribution of $\boldsymbol{Y} \mid \boldsymbol{\theta}, \boldsymbol{d}$ after the data set $\boldsymbol{x}^t$ has been admitted.

Definition 3.4.4. *Call $l(\boldsymbol{\theta} \mid \boldsymbol{x}^t)$ **panel separable** for the panel parameters $\boldsymbol{\theta}_i$, $i \in [m]$, if, given admissible evidence $\boldsymbol{x}^t$, it respects the product form*

$$l(\boldsymbol{\theta} \mid \boldsymbol{x}^t) = \prod_{i \in [m]} l_i(\boldsymbol{\theta}_i \mid \boldsymbol{t}_i(\boldsymbol{x}^t)), \qquad (3.4.5)$$

where $l_i(\boldsymbol{\theta}_i \mid \boldsymbol{t}_i(\boldsymbol{x}^t))$ is a function of $\boldsymbol{\theta}$ only through $\boldsymbol{\theta}_i$ and $\boldsymbol{t}_i(\boldsymbol{x}^t)$ is a statistic of $\boldsymbol{x}^t$, $i \in [m]$.

Note that this likelihood factorisation has common features with the cutting likelihood of Faria and Smith [1997] reviewed in Definition 2.6.4, in the sense that they both separate the complete likelihood into sub-likelihood functions over different components of the vector of interest.

The following theorem then shows that an IDSS is sound a posteriori whenever data admitted into the system has an associated panel separable likelihood.

Theorem 3.4.5. *Assume the IDSS is universally adequate, delegatable, separately informed, cutting and commonly separated at time $t = 0$. If at time $t \in \mathbb{Z}_{\geq 0}$, all data admitted into the system has panel separable likelihood and the joint prior over $\boldsymbol{\theta} \mid \boldsymbol{d}$, $\boldsymbol{d} \in \mathcal{D}$, is absolutely continuous with respect to a Lebesgue measure, then the system is also sound and distributed at time t. Conversely, if at any time t the likelihood is not panel separable over a set of non-zero prior measure over $\boldsymbol{\theta} \mid \boldsymbol{d}$, then the IDSS is no longer sound and distributed.*

The proof of this result can be found in Appendix A.2.2

Importantly, the converse stated in Theorem 3.4.5 implies that the nice properties associated to distributed IDSSs break down whenever Bayesian updating is performed with non panel separable likelihoods. In such a case data needs to be shared between panels and communication channels need to be opened for the IDSS to remain coherent. This for example was shown in the simple binary situation of Section 3.3.4. We briefly mention in Chapter 6 methods that can be employed by an IDSS to deal with non panel separable likelihoods.

Definition 3.4.6. *Say the **separability property** holds for an IDSS if, at any time $t \in \mathbb{Z}_{\geq 0}$, its admissibility protocol only admits data $\boldsymbol{x}^t$ whose associated likelihood is panel separable.*

The following lemma identifies a protocol that selects datasets in such a way that the separability property holds.

Lemma 3.4.7. *Suppose panel independence holds at time $t = 0$ and assume all information presented to the IDSS is the union of I_+^0 and n commonly acknowledged independent sampling schemes. Then, if the admissibility protocol of the IDSS only allows complete data sets from these schemes, there is a unique maximal set of schemes forming I_+^t which ensures soundness and distributivity.*

Proof. Since the likelihood of any set of independent schemes is the product of the likelihoods of the individual schemes, then if the likelihood of any of these schemes is not panel separable over a set of non zero prior measure, the inclusion of the associated data destroys soundness as stated by Theorem 3.4.5. So a maximal amount of sampling evidence that can be admitted into the IDSS is simply the data from the set of schemes whose likelihoods are panel separable. $\square$

Thus, under the conditions of this lemma there is a natural collection of data sets of this type to admit into the system. Note that this is the most informative of all such sound systems in terms of standard measures [O'Hagan and Forster, 2004], for examples ones defined in terms of Kullbach-Leibler information.

Issues are more challenging when information is sparse and data sets in incomplete form might be essential to calibrate the system. This was seen in the various examples of Section 3.3 where data on one or another margin would be admissible but not both. However, we do not deal with this issue here (see Chapter 6 for a discussion).

3.5 Causality in Integrating Systems

3.5.1 A New Definition of Causality

As briefly reviewed in Section 2.3.4, recent works have formalised causal hypotheses to make inferences about the extent of a cause. Here we define a new causal assumption tailored to the needs of an IDSS.

Definition 3.5.1. *Call an IDSS $\{\mathcal{D}, \boldsymbol{d}^0\}$-determined if $\Psi_i^{\boldsymbol{\theta}}$, $i \in [m]$, is a stochastic function of $\Psi_i^{\boldsymbol{\theta}}(L_i, \boldsymbol{d}^0)$, known to G_i, for some prescribed decision $\boldsymbol{d}^0 \in \mathcal{D}$.*

Clearly, if this property is part of the CK-class, then this greatly simplifies the learning each panel needs to undertake. Once they have specified their beliefs about the parameters $\boldsymbol{\theta}_i$ under the decision $\boldsymbol{d}^0$, then the panel beliefs under other decisions can be calculated automatically. Obviously, exactly whether and how this condition is satisfied depends heavily on the domain of the IDSS. Importantly this condition is implicit for classic Perlean CBNs in the following sense.

Lemma 3.5.2. *A CBN with vertex set $\{Y_i : i \in [n]\}$ in a given CK-class is a $\{\mathcal{D}, \boldsymbol{d}^0\}$-determined IDSS whenever $\boldsymbol{d}^0$ is the decision not to intervene but to simply observe the system, $\mathcal{D}$ consists of Perlean interventions, and panel G_i delivers beliefs about $\boldsymbol{\theta}_{B_i}$, $i \in [m]$, where $B_1, \ldots, B_m$ is a partition of $[m]$.*

Proof. This result is a direct consequence of the definition of a CBN given in Definition 2.3.50. Since the decision space consists of Perlean interventions only, each belief $\Psi_i^{\boldsymbol{\theta}}(L_i, \boldsymbol{d}) = \Psi_i^{\boldsymbol{\theta}}(L_i, \mathrm{do}(\boldsymbol{Y}_A = \boldsymbol{y}_A))$, $A \subset [n]$, either coincides with the probabilities in the idle system, coinciding with $A = \emptyset$, or consists of the degenerate distributions having mass one at the manipulated values. Therefore this is a $\{\mathcal{D}, \boldsymbol{d}^0\}$-determined IDSS. $\square$

It can be checked that, for instance, causal hypotheses for CEGs or MDMs provide a $\{\boldsymbol{\mathcal{D}}, \boldsymbol{d}^0\}$-determined IDSS, where $\{\boldsymbol{\mathcal{D}}, \boldsymbol{d}^0\}$ are analogously defined. Within the context of IDSSs it is helpful to extend the usual causal assumptions since the collective might want the flexibility not to map the effects of enacting a Perlean do operation only, but also more complex decision rules, thus embellishing $\boldsymbol{\mathcal{D}}$. Furthermore, the natural comparator $\boldsymbol{d}^0$ for predicting what might happen, might not simply consist of observing the idle system, but rather following routine procedures or past protocols. Therefore our condition provides the basis for more comprehensive analyses which are not simply based on the assumption of the existence of a non-manipulated system.

3.5.2 Admission of Experimental Evidence

Different panels in most IDSSs will want to accommodate not only observational data, in ways discussed in Section 3.4, but also experimental one. Here covariates are controlled and set to certain values and the effect on a response variable is observed. So, for instance, in a nuclear IDSS one module might predict the radiation absorbed by different types of food stuffs when exposed in a controlled and measured way in a laboratory to certain types and durations of radiation. Implicitly, the causal hypothesis is then adopted which assumes that these absorption distributions would be the same were these plants exposed to the same types, durations and levels of exposure on the ground in observational settings. In an actual accident the panel in charge of this absorption module would receive the stochastic process of radiation data from a dispersion/deposition module and match this to laboratory settings at the same values. Essentially, this demands that the parameters in the observational setting can be identified with the parameters in the controlled one.

So most operational IDSSs need to assume causal hypotheses relating controlled experiments and the unfolding disaster. Recall that Daneshkhah and Smith [2004] showed that in the context of BNs the panel independence assumption necessary to ensure an IDSS's distributivity is intimately linked and plausible only when certain causal hypotheses can be entertained.

Here, just as in the observational case, we introduce a class of likelihoods associated to data collected in designed experiments that can allow for distributed Bayesian updating.

Definition 3.5.3. *Call an IDSS $\boldsymbol{e-}$**panel compatible** for a collection of datasets $\boldsymbol{x}^{\boldsymbol{e}} = (\boldsymbol{x}^{e_1}, \ldots, \boldsymbol{x}^{e_p})^{\mathrm{T}}$ obtained from experiments $\boldsymbol{e} = (e_i)_{i \in [p]}^{\mathrm{T}}$ if the likelihood associated to $\boldsymbol{e}$ can be written as*

$$l^{\boldsymbol{e}}(\boldsymbol{\theta} \mid \boldsymbol{x}^{\boldsymbol{e}}) = \prod_{i \in [m]} l_i^e(\boldsymbol{\theta}_i \mid \boldsymbol{t}_i^e(\boldsymbol{x}^{\boldsymbol{e}})),$$

where $l_i^e(\boldsymbol{\theta}_i \mid \boldsymbol{t}_i^e(\boldsymbol{x}^e))$ is a function of the parameter vector $\boldsymbol{\theta}_i$ overseen by panel G_i only, $i \in [m]$, and $t_i^e(\boldsymbol{x}^e)$ is a statistic of $\boldsymbol{x}^e$.

By far the most common suite of such experiments is one composed of collections of independent experiments that can be partitioned across the panels, where each experiment is informative only about the parameters overseen by a particular panel. However a single experiment may be informative for parameters overseen by different panels. If such an experiment were orthogonal over the vectors of parameters under the responsibility of different panels, then the orthogonality would ensure that the likelihood separates across these vector parameters. Therefore, $\boldsymbol{e}$-panel compatibility would still hold in this case.

Again an important special case that automatically implies this separation is when the overarching qualitative structure is a CBN under experimental manipulations.

Lemma 3.5.4. *Suppose an IDSS is $\{\boldsymbol{\mathcal{D}}, \boldsymbol{d}^0\}$-determined, where $\boldsymbol{d}^0 \in \boldsymbol{\mathcal{D}}$ consists of not intervening in the system. Then the IDSS is $\boldsymbol{e}$-panel compatible for a collection of experiments $\boldsymbol{e} = (e_i)_{i \in [p]}^{\mathrm{T}}$, if $\boldsymbol{e}$ consists of independent randomised designed experiments e_k, $k \in [p]$, that observe the response of a variable Y_k in the vertex set of a CBN, where each design consists of the Perlean intervention $\mathrm{do}(\boldsymbol{Y}_{\Pi_k} = \boldsymbol{y}_{\Pi_k})$, for $\boldsymbol{y}_{\Pi_k} \in \boldsymbol{\mathcal{Y}}_{\Pi_k}$.*

Proof. This is a straightforward consequence of the definition of a CBN in Definition 2.3.50. In this setting this implies that the $\boldsymbol{\theta}_k$'s appearing in the manipulated experiment are equal to G_k's corresponding parameters of interest in the observational setting, coinciding to $\boldsymbol{d}^0 \in \boldsymbol{\mathcal{D}}$. But because this is so for the idle control $\boldsymbol{d}^0$, it is also true for all $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$ since the IDSS is $\{\boldsymbol{\mathcal{D}}, \boldsymbol{d}^0\}$-determined. $\square$

Now that we have introduced classes of experiments associated to likelihoods exhibiting the type of separation required for distributivity, we show that Bayesian updating can be distributed to panels in the case the structural consensus includes a CBN.

Theorem 3.5.5. *Suppose the CK-class of a $\{\boldsymbol{\mathcal{D}}, \boldsymbol{d}^0\}$-determined IDSS includes a CBN, where $\boldsymbol{d}^0$ consists of not intervening in the system. Suppose panel independence is in the CK-class and panels oversee separate components of the parameter vector $\boldsymbol{\theta}$. Suppose data admitted in the IDSS at any time t is of the form of both a collection of ancestral random samples of that CBN and a collection of randomised designed experiments as in Lemma 3.5.4. Suppose all these data sets are independent of each other. Then if the IDSS was distributed and sound a priori it remains so at time t.*

Proof. Since all different datasets are independent of each other, the likelihood of the arriving data is simply the product of the likelihoods associated with each component

sample/experiment. So it is sufficient to show that each component likelihood is either panel separable or $\boldsymbol{e}$-panel compatible. Now Lemma 3.5.4 guarantees that the data arising from these experiment is $\boldsymbol{e}$-panel compatible and therefore exhibit the required separability. At the same time Proposition 2.3.13 guarantees that the likelihood associated to the random sample is panel separable. $\qquad\square$

Again analogues of this theorem are also true for a CK-class of an IDSS which contains different overarching qualitative models such as the CEG or the MDM. In particular, Bayesian updating can be distributed to panels whilst retaining the soundness of the IDSS whenever the causal assumptions underpinning the IDSS are such that parameters in the observational setting can be identified with parameters in the controlled one, and the associated likelihoods from these experiments are $\boldsymbol{e}$-panel compatible.

3.6 Examples of Integrating Systems

We saw in Sections 3.4 and 3.5 that, provided an IDSS is such that certain qualitative properties exist over the parameter vectors and the likelihood separates over these parameters in an appropriate way, then the composite system should remain distributed. But how common are such models? The answer is that whilst many systems violate the conditions needed, many others satisfy these. So, by choosing panels appropriately and by demanding that only certain types of unambiguously interpretable data are admitted into the IDSS, it is often possible to build such distributed systems. In this section we present some well known examples when this is possible.

3.6.1 Non Dynamic Case

3.6.1.1 Independence Models and Additive Independence. We begin with a trivial system made up of independent components and a linear utility. Suppose the structural consensus of a CK-class includes the agreement of an independence model over n random variables Y_i, $i \in [n]$ where the density of $\boldsymbol{Y}_i \mid \boldsymbol{\theta}, \boldsymbol{d}$ is parametrised by $\boldsymbol{\theta}_i$, $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$. Suppose panel G_i oversees variables with index in the set B_i, where $B_1, \ldots, B_m$ form a partition of $[n]$. Suppose the utility consensus is such that the utilities considered by the IDSS can only take the form

$$u(\boldsymbol{r}(\boldsymbol{y}, \boldsymbol{d})) = \sum_{i \in [m]} k_i u_i(\boldsymbol{r}_i(\boldsymbol{y}_{B_i}, \boldsymbol{d})),$$

where $\boldsymbol{r}_i$ is a function of $\boldsymbol{d}$ and $\boldsymbol{y}_{B_i}$ only, $i \in [m]$. Then

$$\bar{u}(\boldsymbol{d}) = \sum_{i \in [m]} k_i \bar{u}_i(\boldsymbol{d}),$$

where

$$\bar{u}_i(\boldsymbol{d}) = \int_{\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i} \bar{u}_i(\boldsymbol{d} \mid \boldsymbol{\theta}_i) \pi_i(\boldsymbol{\theta}_i \mid \boldsymbol{d}) \mathrm{d}\boldsymbol{\theta}_i,$$

$\pi_i(\boldsymbol{\theta}_i \mid \boldsymbol{d})$ is G_i's prior over $\boldsymbol{\theta}_i \mid \boldsymbol{d}$ and

$$\bar{u}_i(\boldsymbol{d} \mid \boldsymbol{\theta}_i) \int_{\boldsymbol{\mathcal{Y}}_{B_i}} u_i(\boldsymbol{r}_i(\boldsymbol{y}_{B_i}, \boldsymbol{d})) f(\boldsymbol{y}_{B_i} \mid \boldsymbol{\theta}_{B_i}, \boldsymbol{d}) \mathrm{d}\boldsymbol{y}_{B_i}.$$

This system is clearly a priori distributed. So the SB can devolve her calculations of the expected utility of each $\boldsymbol{d} \in \boldsymbol{\mathcal{D}}$ to the relevant panels. Note that *any* joint prior over $\boldsymbol{\theta} \mid \boldsymbol{d}$ with the given panel margins will clearly give the same function $\bar{u}(\boldsymbol{d})$ and hence the same scores whatever the joint distribution of component panel parameters, so all these potential dependences can be safely ignored as in the example of Section 3.3.2. However, if the utility were not linear, then this would no longer be the case and assumptions associated with panel independence would be needed: see Chapter 5 for an analysis of this issue.

Example 3.6.1. Assume for simplicity that each panel oversees a univariate random variable Y_i and that individually each agreed that $u_i(\boldsymbol{r}(y_i, \boldsymbol{d})) = -y_i^2$. This setting allows us to illustrate a first use of the tower rules of moments of Proposition 2.1.6, which are extensively applied in Chapters 4 and 5. Note that, from Proposition 2.1.5, it holds that

$$\bar{u}_i(\boldsymbol{d}) = \mathbb{E}(Y_i^2 \mid \boldsymbol{d}) = \mathbb{E}(Y_i \mid \boldsymbol{d})^2 + \mathbb{V}(Y_i \mid \boldsymbol{d}).$$

Now, applying the first two tower rules, it follows that

$$\bar{u}_i(\boldsymbol{d}) = \mathbb{E}(\mathbb{E}(Y_i \mid \boldsymbol{\theta}, \boldsymbol{d}))^2 + \mathbb{E}(\mathbb{V}(Y_i \mid \boldsymbol{\theta}, \boldsymbol{d}))^2 + \mathbb{V}(\mathbb{E}(Y_i \mid \boldsymbol{\theta}, \boldsymbol{d})),$$

which is equal to

$$\bar{u}_i(\boldsymbol{d}) = \mathbb{E}(\mu_i \mid \boldsymbol{d})^2 + \mathbb{V}(\mu_i \mid \boldsymbol{d}) + \mathbb{E}(\psi_i \mid \boldsymbol{d}),$$

where μ_i and ψ_i are respectively the mean and the variance of $Y_i \mid \boldsymbol{d}$. Note that in this setting $\bar{u}_i(\boldsymbol{d} \mid \boldsymbol{\theta}) = \mu_i^2 + \psi_i$.

3.6.1.2   Bayesian Networks. We consider BNs as a first example of more complicated multivariate models that can be part of the structural consensus of a CK-class. Two of the results of Section 3.5 - specifically Lemma 3.5.4 and Theorem 3.5.5 - already

considered the use of CBNs in IDSSs. One of the assumptions underlying the validity of both results is that panels oversee disjoint subsets of the vertex set of the CBN. This condition also has to hold for a generic BN in order for the IDSS to provide coherent support.

More specifically, unless each panel G_i oversees a vector $\boldsymbol{Y}_{B_i}$, $i \in [m]$, where $\cup_{i \in [m]} B_i = [n]$ and $\cap_{i \in [m]} B_i = \emptyset$, then soundness and distributivity cannot in general be guaranteed. The proof of Theorem 3.4.3 showed that in order for soundness and distributivity to hold, the panel independence condition needs to be entertained by the IDSS. In the context of BNs this condition corresponds to the independence statement, for every $\boldsymbol{d} \in \mathcal{D}$, $\perp\!\!\!\perp_{i \in [m]} \boldsymbol{\theta}_{B_i} \mid \boldsymbol{d}$. Note that this is a generalisation of the global independence condition introduced in Definition 2.3.9. Under this assumption the following result, extending Proposition 2.3.12 to groups of experts, then can be easily seen to hold.

Proposition 3.6.2. *Suppose the structural consensus of a CK-class includes a BN model with vertex set $\{Y_i : i \in [n]\}$, where panel G_i oversees the vector $\boldsymbol{Y}_{B_i}$, $i \in [m]$, with $B_1, \ldots, B_m$ a partition of $[n]$. Suppose there is a consensus that $\perp\!\!\!\perp_{i \in [m]} \boldsymbol{\theta}_{B_i} \mid \boldsymbol{d}$, $\boldsymbol{d} \in \mathcal{D}$. Then*

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{d}) = \prod_{i \in [m]} \pi_i(\boldsymbol{\theta}_{B_i} \mid \boldsymbol{d}), \tag{3.6.1}$$

$$f(\boldsymbol{y} \mid \boldsymbol{d}) = \prod_{i \in [m]} f_i\left(\boldsymbol{y}_{B_i} \mid \boldsymbol{y}_{\Pi_{B_i}}, \boldsymbol{d}\right), \tag{3.6.2}$$

where $\Pi_{B_i} = \cup_{j \in B_i} \Pi_j$ and

$$f_i\left(\boldsymbol{y}_{B_i} \mid \boldsymbol{y}_{\Pi_{B_i}}, \boldsymbol{d}\right) = \int_{\boldsymbol{\Theta}_{B_i}} f_i(\boldsymbol{y}_{B_i} \mid \boldsymbol{y}_{\Pi_{B_i}}, \boldsymbol{\theta}_{B_i}, \boldsymbol{d}) \pi_i(\boldsymbol{\theta}_{B_i} \mid \boldsymbol{d}) d\boldsymbol{\theta}_{B_i}. \tag{3.6.3}$$

Proof. Equation (3.6.1) follows from the independence of the parameter vectors overseen by different panels, whilst equations (3.6.2) and (3.6.3) are straightforward generalisations of equations (2.3.1) and (2.3.2) in Proposition 2.3.12, respectively, to vectors of parameters associated to disjoint subsets of the vertex set. $\square$

Panel independence therefore guarantees that the marginal distribution of the BN can be written as the product of the panels' marginals just as in the single agent case. Furthermore, learning in BNs where panels oversee disjoint subsets of the vertex set can be distributed to the different groups of experts.

Proposition 3.6.3. *Under the conditions of Proposition 3.6.2, suppose a complete dataset $\boldsymbol{x}$ has been collected from the same population of $\boldsymbol{Y}$. Then the IDSS posterior*

distribution can be written as

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{d}) = \prod_{i \in [m]} \pi_i(\boldsymbol{\theta}_{B_i} \mid \boldsymbol{x}_{Fa_{B_i}}, \boldsymbol{d}), \tag{3.6.4}$$

where $Fa_{B_i} = \cup_{j \in B_i} Fa_j$.

Proof. This follows by noting that equation (3.6.4) can be seen as an instance of equation (2.3.3) in Proposition 2.3.13 applied to vectors of parameters associated to subsets of the vertex set. □

Therefore, under the conditions of Proposition 3.6.3 Bayesian updating can be delegated to the panels and distributivity is assured after the introduction of data. Similar recursions hold in the case the dataset admitted into the IDSS is ancestral.

Example 3.6.4. Consider the network in Figure 1.2 and suppose a BN with this DAG is in the structural consensus of an IDSS. For ease of notation suppress the dependence on $\boldsymbol{d}$. Let

- Y_1: power plant;
- Y_2: source term;
- Y_3: air dispersal;
- Y_4: water dispersal;
- Y_5: deposition;

- Y_6: animal absorption;
- Y_7: human absorption;
- Y_8: human health;
- Y_9: costs;
- Y_{10}: political effects.

Suppose, as denoted by the colors in Figure 1.2 on page 7, that there are six panels G_i, $i \in [6]$, where the sets B_i are such that $B_1 = \{1, 2\}$, $B_2 = \{3, 4, 5\}$, $B_3 = \{6, 7\}$, $B_4 = \{8\}$, $B_5 = \{9\}$ and $B_6 = \{10\}$. Panel independence then implies that

$$\pi(\boldsymbol{\theta}) = \pi_1(\boldsymbol{\theta}_{[2]})\pi_2(\boldsymbol{\theta}_{[5]_2})\pi_3(\boldsymbol{\theta}_{[7]_5})\pi_4(\boldsymbol{\theta}_8)\pi_5(\boldsymbol{\theta}_9)\pi_6(\boldsymbol{\theta}_{10}),$$

$$f(\boldsymbol{y}) = f_1(\boldsymbol{y}_{12})f_2(\boldsymbol{y}_{345} \mid y_2)f_3(y_6, y_7 \mid y_4, y_5)f_4(y_8 \mid y_7)f_5(y_9 \mid y_5, y_6, y_8)f_6(y_{10} \mid y_3, y_9),$$

where $\boldsymbol{y}_{12} = (y_1, y_2)^{\mathrm{T}}$, $\boldsymbol{y}_{345} = (y_3, y_4, y_5)^{\mathrm{T}}$ and $[n]_{i-1} = \{i, \ldots, n\}$.

We can now extend the combination rules of Faria and Smith [1997] (Proposition 2.6.5), based on a group of experts delivering probabilities over the whole network, to our framework, where groups of experts agree to oversee only a subset of the network.

Proposition 3.6.5. *Under the conditions of Proposition 3.6.2, assume the underlying DAG is decomposable, data admitted into the IDSS has panel separable likelihood and each panel agrees to aggregate the density of its members using a conditional logOp. Then, given panel independence holds a priori, the IDSS respects the CEB property if densities are backward sequentially updated.*

Proof. The result follows by first noting that under the assumptions of the theorem, an IDSS so defined is sound and therefore a unique density function can be defined for the whole network. Then noting that panel separable likelihoods in equation (3.4.5) can be thought of as an instance of a cutting likelihood in equation (2.6.1), the proposition follows from Proposition 2.6.5. □

3.6.1.3   Object Oriented Bayesian Networks. Just as for BNs, OOBNs can be used as an integrating tool in IDSSs when panels oversee disjoint subsets of the overall networks. However, in addition, in this case a panel needs to be responsible for all the instantiations of the subnetworks whose density functions are identified. More formally, for a coherent IDSS whose structural consensus include the use of an OOBN, each class of the model has to be under the responsibility of a unique panel, which delivers a unique probability distribution, or a summary of this, for the encapsulated and the output nodes of every object of that class.

Example 3.6.6. Consider again the OOBN defined by the networks in Figures 2.2 and 2.3. In order to guarantee coherency, a unique panel has to oversee the two object of the unique class comprising the variables Y_i' and Y_i'', $i = [6]_1$. Possibly different panels may have jurisdiction over the remaining variables, Y_1 and Y_7.

3.6.1.4   Probabilistic Chain Graphs. PCGs can be included in the structural consensus of coherent and distributed IDSSs just as for BNs. This is because, as noted in Section 2.3.2.5, every CG can be represented as a DAG, whose vertices are the strong components of the initial CG. Suppose a PCG with vertex set $\{Y_i : i \in [n]\}$ has N strong components $\{Y_i : i \in C_i\}$, for $C_i \subseteq [n]$. Suppose further that there are m panels and that G_i oversees $\boldsymbol{Y}_{B_i}$, where $B_i = \cup_{j \in N_i} C_j$ for $N_i \subseteq [N]$, such that $B_1, \ldots, B_m$ form a partition of $[n]$. Then it easily follows that, if $\perp\!\!\!\perp_{i \in [m]} \boldsymbol{\theta}_{B_i} \mid \boldsymbol{d}$, equations (3.6.1)-(3.6.3) hold for PCGs as well. By exploiting the one to one correspondence between a CG and the DAG having as vertices the CG's strong components, it can be shown that adaptations of Propositions 3.6.3 and 3.6.5 hold for PCGs as well.

Example 3.6.7. Consider the PCG in Figure 2.5 and for ease of notation suppress the dependence on $\boldsymbol{d}$. Since this graph has associated DAG in Figure 2.6, under panel

independence and assuming that a panel oversees only one of the three chain components of the DAG, we have that

$$\pi(\boldsymbol{\theta}) = \pi_1(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)\pi_2(\boldsymbol{\theta}_4, \boldsymbol{\theta}_5)\pi_3(\boldsymbol{\theta}_6, \boldsymbol{\theta}_7),$$
$$f(\boldsymbol{y}) = f_1(y_1, y_2, y_3)f_2(y_4, y_5 \mid y_1, y_2, y_3)f_3(y_6, y_7 \mid y_4, y_5).$$

3.6.1.5   Influence Diagrams. IDs can be used within IDSSs to represent in a unique graphical representation structural, utility and policy consensus, since their vertex set includes random, controlled and utility vertices. Importantly, as guaranteed by Lemma 2.5.5, IDs entertains a factorisation of the density function which mirrors the one of BNs. Therefore all the results for IDSSs whose structural consensus includes a BN model translate to generic IDs. This is sufficient to perform distributed inference. However for the purpose of decision making, IDs need an additional constraint. Since joint elicitation of the shape of marginal utility functions between panels are not allowed, then a unique panel must have jurisdiction over all the parents of a utility node. Suppose panel G_i is responsible for $\boldsymbol{Y}_{B_i}$, $i \in [m]$, where $B_i \subset \mathbb{V}$, the indices of the ID's random variables. Then the sets B_i, $i \in [m]$, must be such that, if $\mathbb{U} = \{u_i : i \in [k]\}$, then, for every $j \in [k]$, $P_j \subseteq B_i$ and $P_j \cap B_l = \emptyset$, for every $i \neq l \in [m]$, where P_j is the parent set of u_j.

3.6.1.6   Markov Networks. Whilst the conditions needed for distributivity in the classes of models considered so far are fairly mild and are implied by the assumptions commonly made in practice in single agents domains, these conditions can be fierce in IDSSs whose structural consensus consists of a decomposable MN model can be fierce. From Proposition 2.3.23, a reasonable way to construct panels is in such a way that each is responsible for the marginal distributions over certain collections of cliques of an agreed UG. Now we have a problem because two different panels overseeing adjacent cliques *both* have responsibility for parameters of the margins of variables lying on the separator. Variational independence is therefore compromised and panel independence cannot hold in this setting. Suppose, however, that panels can agree on the same prior over the separators - ensured for example through imposing into the CK-class that the centre's beliefs are strong hyper Markov (see Definition 2.3.21 and Proposition 2.3.22). Even then, if the two adjacent panels update their beliefs autonomously with their own personal information and this information is not shared, then there is no guarantee that the resulting two posterior distributions of the two different panels will remain strong hyper Markov (since Proposition 2.3.23 holds only in the case of complete datasets).

One simple solution to this problem when the underlying UG is decomposable is to give precedence to one panel's information about a particular separator and ignore all others. This is equivalent to selecting a CG representation of this decomposable UG where
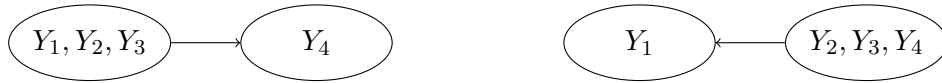
FIGURE 3.3: Possible conversions of the undirected graph in Figure 2.4 into a directed acyclic graph.

responsibility for a particular separator is delegated entirely to the panel delivering the parent clique probabilities. The other panel is then only responsible for delivering the probability judgements about its clique probabilities conditional on the values of the separator. However with no 'causal directionality' implied by the graph, the choice of the responsible panel looks rather arbitrary. Furthermore this type of protocol does not extend straightforwardly to generic MN models and destroys the symmetry of the original agreed model. It is interesting to note that an analogous situation is sometimes encountered in meta analyses where experiments have been performed by different panels and posterior results on these overlapping margins are communicated [see e.g. Jirousek and Vejnarova, 2003, Massa and Lauritzen, 2010].

Example 3.6.8. Consider the UG in Figure 2.4, having two cliques $\{Y_1, Y_2, Y_3\}$ and $\{Y_2, Y_3, Y_4\}$, overseen by G_1 and G_2 respectively. This UG can be converted both into the DAG on the left of Figure 3.3, where panel G_1 has jurisdiction over the separator, and the one on the right, where conversely G_2 is responsible for the separator.

We note here however that there is situation in which updating can be distributed across panels using the original MN model that does not break the coherence of the IDSS (which recall does not respect variational independence). We formalise this case in the following theorem.

Theorem 3.6.9. *Let $\mathcal{G}$ be an decomposable UG, $\mathcal{C}$ and $\mathcal{S}$ the sets of cliques and separators of $\mathcal{G}$ respectively, and assume panel G_i oversees the variables in the clique $C_i \in \mathcal{C}$. Assume the IDSS is sound a priori and that each panel G_i plans to update its beliefs using the dataset $\boldsymbol{x}_i^t$, independent to the ones of other panels. Then the IDSS is also sound a posteriori iff $\boldsymbol{t}_i(\boldsymbol{x}_{i,S}^t)$, the sufficient statistic for $\boldsymbol{\theta}_S$ of $\boldsymbol{x}_i^t$, is equal for every $S \in \mathcal{S}$ and for every $C_i \in \mathcal{C}$ such that $S \subset C_i$.*

The proof of this result is reported in Appendix A.2.3.

Importantly this result shows that strong hyper Markov distributions are preserved when updated with incomplete datasets of a particular type - i.e. partial datasets covering together the whole graph and having consistent sufficient statistics over the separators.

Example 3.6.10. Consider again the MN with graph in Figure 2.4 and suppose a covariance selection model is defined over this network as shown in Section 2.3.2.4.

Suppose two panels oversee the two cliques of this graph and assume that a strong hyper Markov distribution has been agreed over the graph. Suppose panels collect random samples from the same family of $\boldsymbol{Y}$ only over the variables they oversee, denoted as $\boldsymbol{x}^1 = (\boldsymbol{x}_1^{1\mathrm{T}}, \boldsymbol{x}_2^{1\mathrm{T}}, \boldsymbol{x}_3^{1\mathrm{T}})^{\mathrm{T}}$ for G_1 and $\boldsymbol{x}^2 = (\boldsymbol{x}_2^{2\mathrm{T}}, \boldsymbol{x}_3^{2\mathrm{T}}, \boldsymbol{x}_4^{2\mathrm{T}})^{\mathrm{T}}$ for G_2. Let a generic $\boldsymbol{x}_j^i = (x_{jk}^i)_{k \in [n_i]}^{\mathrm{T}}$. Then, as shown in equation (2.3.5), coherence is guaranteed a posteriori iff $\sum_{i \in [n_1]}(x_{2i}^1)^2 = \sum_{i \in [n_2]}(x_{2i}^2)^2$, $\sum_{i \in [n_1]}(x_{3i}^1)^2 = \sum_{i \in [n_2]}(x_{3i}^2)^2$ and $\sum_{i \in [n_1]} x_{2i}^1 x_{3i}^1 = \sum_{i \in [n_2]} x_{2i}^2 x_{3i}^2$.

3.6.1.7 Staged Trees and Chain Event Graphs. Suppose for each possible decision $\boldsymbol{d} \in \mathcal{D}$, panels can agree the topology of the underlying event tree $\mathcal{T}(\boldsymbol{d})$ and suppress the dependence on the decisions for ease of notation. Suppose also all agree that panel G_i, $i \in [m]$, should deliver the edge probability vectors $\boldsymbol{\theta}_{ij}$ associated with edges emanating from each non-leaf vertex v_{ij}, $j \in [m_i]$, $m_i \in \mathbb{Z}_{\geq 1}$. Let $\boldsymbol{\theta}_{ij} = (\theta_{ijk})_{k \in [m_{ij}]}^{\mathrm{T}}$, for an $m_{ij} \in \mathbb{Z}_{\geq 1}$, $\boldsymbol{\theta}_i^{\mathrm{T}} = (\boldsymbol{\theta}_{ij})_{j \in [m_i]}^{\mathrm{T}}$ and $\boldsymbol{\theta}^{\mathrm{T}} = (\boldsymbol{\theta}_i^{\mathrm{T}})_{i \in [m]}$. Then, for example, under random sampling

$$l(\boldsymbol{\theta} \mid \boldsymbol{x}) = \prod_{i \in [m]} l_i(\boldsymbol{\theta}_i \mid \boldsymbol{x}), \tag{3.6.5}$$

where

$$l_i(\boldsymbol{\theta}_i \mid \boldsymbol{x}) = \prod_{j \in [m_i]} l_{ij}(\boldsymbol{\theta}_{ij} \mid \boldsymbol{x}), \tag{3.6.6}$$

and

$$l_{ij}(\boldsymbol{\theta}_i \mid \boldsymbol{x}) = \prod_{k \in [m_{ij}]} \theta_{ijk}^{x_{ijk}}, \tag{3.6.7}$$

for $\sum_{k \in [m_{ij}]} \theta_{ijk} = 1$ and x_{ijk} is the number of units in the sample $\boldsymbol{x}$ reaching vertex v_{ij} and then proceeding down the $k-$th edge. Clearly this likelihood separates over the parameter vectors overseen by different panels. It can be similarly shown that this separability holds if the dataset $\boldsymbol{x}$ is ancestral - see Smith [2010].

The conditions for distributivity in IDSSs whose structural consensus includes either a staged tree or a CEG are slightly stronger, but often met in practice. Let $\mathcal{T}$ be a staged tree now, with stage set $\mathbb{W} = \{w_i : i \in [n]\}$. Let $\boldsymbol{\theta}_i^{\mathbb{W}}$ be the probability vector associated to a stage w_i, $i \in [n]$. Unless $\boldsymbol{\theta}_i^{\mathbb{W}} = \boldsymbol{\theta}_{ij}$ for every $j \in [m_i]$ but one only $i \in [m]$, the distributed learning of equations (3.6.5)-(3.6.7) breaks down, since different panels have jurisdiction over parameters that are identified (just as for OOBNs). However, panels can be constituted in such a way that this condition is met. Note that distributivity also holds whenever panels oversee disjoint subsets of the position set (or equivalently disjoint subsets of the vertex set of a CEG), since the position set is a finer partition of the situations of a tree than the stage set.

Thus CEGs and staged trees can provide a framework where data can be quickly accommodated in a distributed way (using the Dirichlet-Multinomial recursions discussed in Section 2.3.2.7), given that the panels oversee either the positions or the stages of a CEG/staged tree.

Example 3.6.11. Consider the staged tree in Figure 2.7. For distributivity to hold, a unique panel needs to have jurisdiction over the positions v_3 and v_4 since these are in the same stage.

3.6.2 Dynamic Models

The necessary separation conditions for IDSSs are not only entertained by non-dynamic models, as exemplified in Section 3.6.1 in a variety of domains, but also in situations where probabilities are allowed to be recursively updated in a dynamic fashion. For ease of notation, in this section we leave implicit the dependence on the decisions $\boldsymbol{d}$.

3.6.2.1 Multiregression Dynamic Models. The first model class we consider is the MDM. Since each time slice of an MDM conditionally on the past is described by a DAG, which does not change through time, the results concerning the use of BNs in IDSSs transfer to these dynamic models straightforwardly.

We note that since a dynamic version of global independence is always assumed in MDMs (Proposition 2.3.39), then panel independence is guaranteed whenever panels oversee disjoint subsets of the vertex set. Furthermore, the densities $f(\boldsymbol{y}_t \mid \boldsymbol{y}^{t-1})$ can be written as the product of functions overseen by individual panels. We formalise these concepts in the following proposition

Proposition 3.6.12. *Suppose the structural consensus of a CK-class includes an MDM model whose associated DAG has vertex set $\{\boldsymbol{Y}_i^T : i \in [n]\}$, where panel G_i oversees the vector $\boldsymbol{Y}_{B_i}^T$, $i \in [m]$, where $B_1, \ldots, B_m$ form a partition of $[n]$. It then follows that for every $t \in T$,*

$$\pi(\boldsymbol{\theta}(t) \mid I^{t-1}) = \prod_{i \in [m]} \pi_i(\boldsymbol{\theta}_{B_i}(t) \mid I^{t-1}), \tag{3.6.8}$$

$$f(\boldsymbol{y}(t) \mid \boldsymbol{y}^{t-1}) = \prod_{i \in [m]} \int_{\boldsymbol{\Theta}_{B_i}} f_i(\boldsymbol{y}_{B_i}(t) \mid \boldsymbol{y}_{\Pi_{B_i}}^t, \boldsymbol{y}_{B_i}^{t-1}, \boldsymbol{\theta}_{B_i}) \pi_i(\boldsymbol{\theta}_{B_i}(t) \mid I^{t-1}) \mathrm{d}\boldsymbol{\theta}_{B_i}(t). \tag{3.6.9}$$

Proof. Equation (3.6.8) is implied by the independence structure underlying MDMs specified in Proposition 2.3.39. Equation (3.6.9) is an instance of equations (2.3.10) and (2.3.11) in Proposition 2.3.40 applied to vectors of parameters overseen by different panels. $\square$

Example 3.6.13. Consider the MDM whose associated DAG is in Figure 2.11 and suppose panels G_1 and G_2 have jurisdiction over $(\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T)$ and $(\boldsymbol{Y}_3^T, \boldsymbol{Y}_4^T)$, respectively. Panel independence in this setting implies that, for any $t \in [T]$, $(\boldsymbol{\theta}_1(t), \boldsymbol{\theta}_2(t)) \perp\!\!\!\perp (\boldsymbol{\theta}_3(t), \boldsymbol{\theta}_4(t)) \mid I^{t-1}$. Furthermore the marginal density of the observables, conditional on the past, can be written as

$$f(\boldsymbol{y}(t) \mid I^{t-1}) = f_2(y_3(t), y_4(t) \mid y_2(t), I^{t-1}) f_1(y_1(t), y_2(t) \mid I^{t-1}),$$

where

$$f_2(y_3(t), y_4(t) \mid y_2(t), I^{t-1}) = \int_{\boldsymbol{\Theta}_{34}} f_2(y_3(t), y_4(t) \mid y_2(t), I^{t-1}, \boldsymbol{\theta}_{34}) \pi_2(\boldsymbol{\theta}_{34} \mid I^{t-1}) \mathrm{d}\boldsymbol{\theta}_{34},$$

$$f_1(y_1(t), y_2(t) \mid I^{t-1}) = \int_{\boldsymbol{\Theta}_{12}} f_1(y_1(t), y_2(t) \mid I^{t-1}, \boldsymbol{\theta}_{12}) \pi_1(\boldsymbol{\theta}_{12} \mid I^{t-1}) \mathrm{d}\boldsymbol{\theta}_{12}.$$

We are also able to show that the combination rules of Faria and Smith [1997], generalised to IDSSs in Proposition 3.6.5, can be extended to dynamic frameworks.

Proposition 3.6.14. *Under the conditions of Proposition 3.6.12, assume the underlying DAG is decomposable and each panel agrees to aggregate the density of each member using a conditional logOp. Then the IDSS respects the CEB property if densities are backward sequentially updated through the vertices of the DAG and through time.*

Proof. The proposition follows by Proposition 3.6.5, since each time slice of the MDM can be thought of, conditional on the past, as a BN, whose underlying DAG does not change through time. □

3.6.2.2 Dynamic Chain Graphs. A generalisation of the class of MDM models is the DCG model class introduced in Section 2.3.3.3. As discussed for PCG models, the CG associated to a DCGs can be converted into a DAG whose vertex set consists of the strong components of the initial CG. Therefore, as for PCGs and BNs, the conditions required for DCGs to be in the structural consensus of a coherent IDSS are the same as the ones of MDMs, but applied to the strong components of the associated CG.

3.7 Conclusions

In this chapter we have introduced the IDSS framework to coherently aggregate the judgements of separate panels of experts agreeing on a common knowledge base. This extends standard Bayesian reasoning in single agent frameworks to ones where beliefs are delivered by many experts. We have shown in Theorem 3.4.3 that the conditions

for the existence of a coherent IDSS are not overly restrictive. By introducing a new notion of causality, we have then demonstrated that Bayesian learning can be distributed to the different panels of experts, guaranteeing fast inferential routines, just as in the standard Bayesian framework. We have then concluded the chapter showing that a suite of standard Bayesian graphical models can be part of the structural consensus of an IDSS. To do this, we have generalised standard independence notions and learning routines of these common graphical models.

Chapter 4

Distributed Propagation in Integrating Systems

The IDSS methodology introduced in Chapter 3 provides a coherent framework for combining expert judgements and models into a unique entity. Inference can be also distributed in such systems as extensively illustrated in the previous chapter. We now focus on methods to exactly compute the expected utility scores of the available decisions to support decision centres in policy choices. Some of these results have been already reported in Leonelli and Smith [2015].

A commonly used naive method is for the group to simply plug in estimates of observables to compare the efficacy of different policies. However, in general, a proper risk analysis of the composite must also fold in measures of uncertainty about the outputs different panels deliver to the IDSS. It has been known for some time that, even in very simple scenarios, ignoring uncertainties can lead a decision centre into choosing the wrong course of action [see e.g. Leonelli and Smith, 2013a, and the examples of Section 4.4]. This is because expected utility scores for competing suites of countermeasures often formally depend on these uncertainties.

In this chapter we consider a rather large CK-class, describing a new family of highly asymmetric and dynamic decision problems with partial utility independence structures, and deduce conditions that ensure expected utilities can be written as functions of the panels' delivered beliefs only. This can be done by introducing five structural assumptions specifying the group's agreement in this setting. To our knowledge this is the first application of partial utility specifications to the computation of expected utilities in graphical decision models. We then develop distributed propagation algorithms over the network of Expert Systems (ESs) of the IDSS between the panels and the SB. We are able to demonstrate that it is surprisingly simple to calculate propagation algorithms

based on the autonomous panels' computations using standard backward induction, albeit in this novel and potentially very complex setting. These recursions extend standard evaluation algorithms to dynamic asymmetric domains with both probabilities and utilities individually delivered by groups of experts. Distributivity conditions imply that the IDSS is able to quickly produce forecasts and expected utility scores, enabling users to interrogate the system in real time.

For ease of notation, in this chapter we suppress the dependence on the parameter vectors and develop algorithms for the marginal distribution over the observables only. We have extensively studied in Chapter 3 conditions over the parameters that can guarantee that the marginal distribution over the whole IDSS can be written as the product of the ones associated to each individual panel. Specifically, this is the case whenever panel independence holds. We explicitly label again the dependence on the parameter vectors in the examples of Section 4.4.

The chapter is organised as follows. Section 4.1 introduces the five structural assumptions defining a CK-class comprising many of the models commonly used in practice. In Section 4.2 we show that these assumptions correspond to an instance of the common knowledge axioms introduced in Chapter 3. In Section 4.3 we define our distributed propagation algorithms to compute various expected utility scores. Section 4.4 presents the workings of the algorithms for a variety of examples [that can also be found in Leonelli and Smith, 2013a,b, 2015, Smith et al., 2015]. These examples also demonstrate the dangers associated to a non complete uncertainty handling due to the lack of integration of the modules. We conclude with a discussion.

4.1 An Instance of an Asymmetric and Dynamic Integrating System

4.1.1 The Common Knowledge Class

Let $\{\boldsymbol{Y}(t)\}_{t \in [T]}$, $[T] = \{1, 2, \ldots, T\}$, be a multivariate time series with finite horizon T partitioned into n multivariate time series $\{\boldsymbol{Y}_i(t)\}_{t \in [T]}$, with $i \in [n]$. Each individual time series $\{\boldsymbol{Y}_i(t)\}_{t \in [T]}$ is overseen by the panel of experts G_i and includes all the variables associated to the i-th DSS. The sample space of $\boldsymbol{Y}(t)$ is $\boldsymbol{\mathcal{Y}} = \times_{i \in [n]} \boldsymbol{\mathcal{Y}}_i$, where $\boldsymbol{\mathcal{Y}}_i$ is the sample space of $\boldsymbol{Y}_i(t)$, $i \in [n]$, $t \in [T]$. We denote with lower case letters instantiations of these random vectors and, for any $t \in [T-1]$, the sample space of $(Y(t), Y(t+1))$ is $\boldsymbol{\mathcal{Y}} \times \boldsymbol{\mathcal{Y}}$. Recall that $\boldsymbol{Y}^t = (\boldsymbol{Y}(1)^{\mathrm{T}}, \ldots, \boldsymbol{Y}(t)^{\mathrm{T}})^{\mathrm{T}}$ and let $\Pi_i \subseteq [i-1]$. Just as in Chapter 3 we assume here that the collective is jointly responsible for the definition of the necessary overarching probabilistic, preferential and decision structures. The structure within each
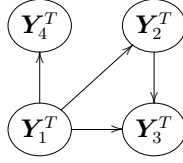
FIGURE 4.1: Example of a directed acyclic graph of a distributed dynamic model depicting relationships between processes, not variables.

individual module is on the other hand agreed by the members of the relevant panel only [as often in practice, see e.g. Von Winterfeldt and Edwards, 1986]. For ease of notation in Section 4.1.1.1 we suppress the dependence on the decisions $\boldsymbol{d}$ with no loss of generality. We again explicitly label this dependence from Section 4.1.1.2.

4.1.1.1 Distributed Dynamic Models. The overall statistical model the collective needs to agree upon is, for the purpose of this chapter, a new dynamic graphical Bayesian model customised to the needs of multi-ESs, here called **Distributed Dynamic Model (DDM)**. In a DDM, just as for MDMs and DCGs, relationships between time series are depicted by a DAG whose vertices are $\boldsymbol{Y}_i^T$, $i \in [n]$. Here we assume that the vector $\boldsymbol{Y}^T$ includes all the variables the collective is planning to take into account during the analysis.

We are now ready to formally define the DDM model class.

Definition 4.1.1. *A* DDM *for the time series* $\{\boldsymbol{Y}(t)\}_{t \in [T]}$ *consists of:*

- $n-1$ *conditional independence statements for each time point* $t \in [T]$ *of the form*

$$\boldsymbol{Y}_i(t) \perp\!\!\!\perp \boldsymbol{Y}_{[i-1]\backslash\Pi_i}^t \mid \boldsymbol{Y}_{\Pi_i}^t, \boldsymbol{Y}_i^{t-1}; \tag{4.1.1}$$

- *a DAG* $\mathcal{G}$ *with vertex set* $V(\mathcal{G}) = \{\boldsymbol{Y}_i^T : i \in [n]\}$ *and edge set* $E(\mathcal{G})$ *including an element* $(\boldsymbol{Y}_j^T, \boldsymbol{Y}_i^T)$ *if* $j \in \Pi_i$, $i \in [n]$.

The conditional independence structure of the DDM implies that the only information to infer $\boldsymbol{Y}_i(t)$ from $\boldsymbol{Y}_{[i-1]\backslash\Pi_i}^t$, $\boldsymbol{Y}_{\Pi_i}^t$ and $\boldsymbol{Y}_i^{t-1}$ is from $\boldsymbol{Y}_{\Pi_i}^t$ and $\boldsymbol{Y}_i^{t-1}$.

An example of a DAG associated to a DDM is presented in Figure 4.1, corresponding to a dynamic variant of the BN in Figure 2.1. Such a DAG, in contrast to the more common BN whose DAG represents relationships between single variables, specifies relationships across the components of different multivariate time series. It is important to note once more that statements embodied within this DAG are *qualitative* in nature and so in particular can more easily provide the framework for a CK-class [Smith, 1996].

Since the vertices of the underlying graph are time series, the topology of the DAG does not change through time. Therefore, each time slice $\boldsymbol{Y}(t)$ of a DDM, conditionally on the past, can also be described graphically by a DAG with the same topology. This topology remains constant as time progresses. On the other hand the associated probabilities are allowed to be dynamically updated through time, using for example the DLM approach illustrated in Section 2.3.3.1. Each time slice DAG has vertex set equal to $\{\boldsymbol{Y}_i(t) : i \in [n]\}$ and edges $(\boldsymbol{Y}_j(t), \boldsymbol{Y}_i(t))$ if $j \in \Pi_i$, $i, j \in [n]$. We call this DAG the **time slice DAG** of the DDM.

Just as for other Bayesian graphical models in the literature, the DDM can be associated with a factorisation of the probability density function, which depends on the topology of the associated DAG. Specifically, as a direct consequence of the conditional independence structure associated with a DDM, we have the following result.

Proposition 4.1.2. *The joint probability density function f associated to a DDM for the time series $\{\boldsymbol{Y}(t)\}_{t \in [T]}$ can be written as*

$$f\left(\boldsymbol{y}^T\right) = \prod_{t \in [T]} \prod_{i \in [n]} f_{t,i}\left(\boldsymbol{y}_i(t) \mid \boldsymbol{y}_{\Pi_i}^t, \boldsymbol{y}_i^{t-1}\right).$$

Proof. The result follows by first applying the chain rule of probabilities so that

$$f(\boldsymbol{y}^T) = \prod_{t \in [T]} \prod_{i \in [n]} f_{t,i}(\boldsymbol{y}_i(t) \mid \boldsymbol{y}_1(t), \dots, \boldsymbol{y}_{i-1}(t), \boldsymbol{y}^{t-1}),$$

and then applying the conditional independences in equation (4.1.1). $\qquad \square$

For the purpose of the collective specification of the overarching probability model, it is only relevant that the probability density can be *qualitatively* written as a product of the terms $f_{t,i}$. The actual algebraic form of these terms and the *quantitative* specification of the associated parameters is agreed, as we specify below and as usual in practice, by the members of the relevant panel only. In this sense the algorithms we derive in Section 4.3 are built on the agreed *qualitative* framework within the collective CK-class and are driven by the topology of the agreed DAG.

We note here that the DDM model class is very large. Particular instances of the DDM have been extensively studied in the literature. For example, from Proposition 2.3.40 it follows that the MDM model class, and therefore also the LMDM one, is a particular instance of a DDM. Proposition 2.3.12 guarantees that BNs with appropriate global independence conditions are a member of the DDM class. It can be similarly shown that PCGs and DCG (Sections 2.3.2.5 and 2.3.3.3 respectively) can be thought of as instances of a DDM under certain assumptions. Consequently, all these models

can be used to embellish the qualitative structure of a DDM with explicit probabilistic specifications.

We are now ready to make the following assumption. Recall that A_i' is the index set of the non-descendants of $\boldsymbol{Y}_i^T$, $i \in [n]$.

Structural Assumption 4.1.3. *The collective agrees to:*

- *describe the predictive factorisation of $\boldsymbol{Y}^T$ by a DDM, whose DAG is connected and decomposable;*

- *assume that the elements of $\{\boldsymbol{Y}(t)\}_{t \in [T]}$ are observed according to the order defined by the following rules:*

  - *$\boldsymbol{Y}_i(t_1)$ is observed before $\boldsymbol{Y}_j(t_2)$ if $t_1 < t_2$, for $i, j \in [n]$;*
  - *$\boldsymbol{Y}_j(t)$ is observed before $\boldsymbol{Y}_i(t)$ if $j \in A_i'$.*

The requirement that the graph is decomposable is simply a technical one, similar to those used in junction trees propagation algorithms [see e.g. Lauritzen, 1992, 1996, Smith, 2010]. This condition provides the basis for fast computational algorithms for BNs. In particular this ensures that no new dependencies are introduced in the IDSS through the backward induction steps we define below. Note that any DAG can be converted into a decomposable one which then gives a valid (albeit inefficient) representation of the underlying processes [see e.g. Smith, 2010, for an explicit description of this embedding, and Appendix B]. In this sense this assumption is not too fierce. Furthermore, we can assume without any loss of generality that the network is connected, since if this were not the case, then, when the structural assumptions we introduce below hold, the overall problem could be decomposed into smaller and independent ones that could be treated separately.

However, more critical is the assumption, as expressed in the second bullet, that it is possible to observe all the quantities the collective planned to observe in the order they happen. It has long been known that when the delivery of some of the data is delayed, the underlying conditional independence structure associated to certain instances of a DDM breaks down [Queen and Smith, 1993] and previously uncorrelated time series overseen by different panels could then become highly correlated, thus destroying the distributivity of the system and the validity of the propagation algorithms. For the purpose of this chapter we assume that the receipt of information is never delayed. We briefly discuss two potential practical ways of addressing violations of this assumption in Chapter 6.

4.1.1.2   An Asymmetric Decision Space. As we specify below, the structure of the decision space the collective shares assumes that a potential decision centre has the possibility of intervening after having observed any variable in the system. For $i \in [n]$ and $t \in [T]$, let $\mathcal{D}_i(t)$ be the decision space available after having observed $\boldsymbol{Y}_i(t)$, and $\mathcal{D}(0)$ be the decision space associated to an initial decision. We also let $\boldsymbol{\mathcal{D}}_B(t) = (\mathcal{D}_i(t))_{i \in B}^{\mathrm{T}}$, $\boldsymbol{\mathcal{D}}_B^t = (\mathcal{D}(0), \boldsymbol{\mathcal{D}}_B(1), \ldots, \boldsymbol{\mathcal{D}}_B(t))^{\mathrm{T}}$ and $\boldsymbol{\mathcal{D}}^t = \boldsymbol{\mathcal{D}}_{[n]}^t$, for $B \subseteq [n]$. We also denote with $\boldsymbol{d}_B(t)$ and $\boldsymbol{d}_B^t$ generic elements of $\boldsymbol{\mathcal{D}}_B(t)$ and $\boldsymbol{\mathcal{D}}_B^t$ respectively.

We next make the following assumption. Recall that A_i is the set of the indices of the vectors in the ancestral set of $\boldsymbol{Y}_i^T$, $i \in [n]$.

Structural Assumption 4.1.4. *The collective agrees:*

- *the specification of the decision spaces $\mathcal{D}(0)$ and $\mathcal{D}_i(t)$, $i \in [n]$, $t \in [T]$, defining the acts a decision centre might take;*

- *to assume that the choice of a decision $d_i(t_2) \in \mathcal{D}_i(t_2)$ is not constrained by a decision $d_j(t_1) \in \mathcal{D}_j(t_1)$ if $j \notin A_i'$ in the DAG of the DDM, $j < i$, $t_1 \leq t_2$;*

- *to commit to a decision $d_i(t) \in \mathcal{D}_i(t)$ only after having observed the value of $\boldsymbol{Y}_{A_i}(t)$ and $\boldsymbol{Y}^{t-1}$, and having already committed to decisions $\boldsymbol{d}_{A_i}(t)$ and $\boldsymbol{d}^{t-1}$;*

- *that the underlying DDM remains valid under any policy choice open to the centre.*

Structural Assumption 4.1.4 guarantees that the graphical framework of the IDSS remains unaffected after a decision is taken, so that the system provides a coherent picture of the problem throughout the unfolding of events and actions. This is because under the assumption above the topology of the time slice DAGs does not change. So the algorithms we define in Section 4.3 are still able to compute coherent expected utility scores through propagation. Of course we can still allow for the possibility that the probability judgements *within* that structure might change in response to a decision - they usually do.

Example 4.1.5. To illustrate Structural Assumption 4.1.4 we consider the diagram of the time slice DAG at time t of our network, reported in Figure 4.2, which includes four decision spaces $\mathcal{D}_i(t)$, $i \in [4]$. Note that this is not a simple ID in reduced form since Structural Assumption 4.1.4 does not guarantee that the decision spaces are totally ordered. In fact these decision spaces only need to be partially ordered consistently with the DAG of the DDM. Therefore, for instance, there is no fixed order in which a decision centre commits to decisions $d_2(t) \in \mathcal{D}_2(t)$ and $d_4(t) \in \mathcal{D}_4(t)$. The constraints associated with this partial order are denoted in Figure 4.2 by the absence of an edge between these two decision spaces. A decision centre needs to commit to one of these
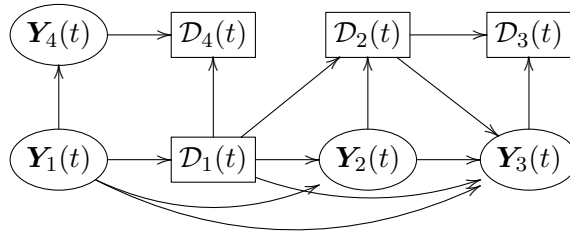
FIGURE 4.2: A time slice directed acyclic graph of the distributed dynamic model in Figure 4.1 including decision nodes.

decisions, $d_i(t) \in \mathcal{D}_i(t)$ say, only after having observed the value of $\boldsymbol{Y}_i(t)$, $i \in [4]$: in our notation only after having observed $\boldsymbol{Y}_{A_i}(t)$ and $\boldsymbol{Y}^{t-1}$ as specified by the third bullet of Structural Assumption 4.1.4. Of course a decision $d_i(t) \in \mathcal{D}_i(t)$ is made after having already committed to $\boldsymbol{d}^{t-1} \in \boldsymbol{\mathcal{D}}^{t-1}$. We further assume that the overall decision space is such that $\mathcal{D}_4(t) \times \mathcal{D}_2(t)$, so that in particular these two decision spaces do not constrain one another. In general decision spaces that are not connected by an edge in the non-reduced representation of the network in Figure 4.2 cannot be mutually constrained.

The second assumption about the structure of the decision problem concerns a set of irrelevance statements.

Structural Assumption 4.1.6. *The collective agrees that*

$$f_{t,i}\left(\boldsymbol{y}_i(t) \mid \boldsymbol{d}^T, \boldsymbol{y}_{\Pi_i}^t, \boldsymbol{y}_i^{t-1}\right) = f_{t,i}\left(\boldsymbol{y}_i(t) \mid \boldsymbol{d}_{A_i'}^t, \boldsymbol{y}_{\Pi_i}^t, \boldsymbol{y}_i^{t-1}\right), \qquad (4.1.2)$$

for $i \in [n]$ and $t \in [T]$.

Equation (4.1.2) states that a random vector $\boldsymbol{Y}_i(t)$ does not functionally depend on the decisions that are not included in $\boldsymbol{\mathcal{D}}_{A_i'}^t$. Alternatively this structural assumption could be represented as a set of extended conditional independences [Dawid and Constantinou, 2014]. This assumption is a very weak one. For example the *sufficiency theorem* of Smith [1989a,b] guarantees that a decision centre can always find *one* Bayes optimal decision based on a decision rule which respects these statements. We further note here that within each time slice this assumption is an instance of the *causal consistency lemma* of Cowell et al. [1999], but applied to this more general setting. The lemma guarantees that decisions can have a direct influence only on variables that are yet to be observed. More generally here Structural Assumption 4.1.6 implies the lemma holds for partially ordered decisions and decision spaces that are not simply product spaces.

Example 4.1.7. Consider the setting of Example 4.1.5. Since there is no fixed order between $\boldsymbol{Y}_2(t)$ and $\boldsymbol{Y}_4(t)$, Structural Assumption 4.1.6 demands that $\boldsymbol{Y}_4(t)$ does not functionally depend on $d_2(t) \in \mathcal{D}_2(t)$. Similarly, we require that $\boldsymbol{Y}_2(t)$ does not functionally

depend on $d_4(t) \in \mathcal{D}_4(t)$. This can be noted in the diagram of Figure 4.2 by the absence of edges between these nodes.

These are the irrelevances the collective *needs* to be ready to assume. Of course they might believe that some decisions do not have any direct effect to additional variables and thus assume further irrelevances. For example they might believe that the initial decision space $\mathcal{D}(0)$ is irrelevant for the outcomes of the variables at the first time point, $\boldsymbol{Y}(1)$. Such additional assumptions do not affect the validity of our algorithms.

4.1.1.3 Compatible Utility Factorisations. The last overarching agreement the collective needs to find concerns the utility factorisation. We suppose the time series with index in $U \subseteq [n]$ to be the attributes of the decision problem. For $i \in U$, define $\boldsymbol{r}_i = \boldsymbol{r}_i \left(\boldsymbol{y}_i^T, \boldsymbol{d}_{A_i}^T \right)$ to be a function of both $\boldsymbol{y}_i^T$ and $\boldsymbol{d}_{A_i}^T$ and let $\boldsymbol{r}_B^{\mathrm{T}} = (\boldsymbol{r}_i)_{i \in B}$ for $B \subseteq [n]$. Note that each vertex $\boldsymbol{Y}_i^T$ of the DAG of the DDM, for $i \in U$, can be uniquely associated with an $\boldsymbol{r}_i$. We show below that this notation is very concise in depicting utility independent statements between time series under the responsibility of different panels. For simplicity we assume that $i \in U$ for all $i \in Le$, where Le is the index set of the leaves of the DAG, since otherwise the associated variables could be simply deleted without affecting the results of the analysis.

In the multi-expert setting we study here joint utility elicitations across different panels in a single integrating decision conference are only rarely possible [see e.g. Chapter 11 of French et al., 2009, and our review in Section 2.6.1]. So for example it is typically possible to elicit the scores associated with the overall weight of one attribute over another, for example as expressed by the *criterion weights* of multiattribute independent utilities. But other more detailed elicitations, for example the appropriate forms of the marginal utility functions, are better delegated to those closest to understanding the consequences of such attributes [for an illustration of why this is so, see Von Winterfeldt and Edwards, 1986]. However, for this type of delegation to be formally justified, it is first necessary to assume that the collective is prepared to entertain certain sets of preferential independences in order to be able to elicit, through individual panels' assessments, a joint utility function.

Here we define a new multiattribute utility factorisation compatible with the DAG of the multi-expert DDM we introduced above. Specifically, this first assumes that the time series with index in U belonging to different ancestral components of the DAG of the DDM are GAI (see Definition 2.4.10). The utility independence structure within each of these components is then assumed to be described by a member of a certain class of *directional utility diagrams* introduced in Definition 2.4.25.

Definition 4.1.8. *Let $\mathcal{G}$ be the DAG of a DDM of a time series $\{\boldsymbol{Y}(t)\}_{t\in[T]}$ and relabel a vertex $\boldsymbol{Y}_i^T$ of $\mathcal{G}$ with $\boldsymbol{r}_i$, $i \in [n]$. Let $U \subseteq [n]$ be the index set of the attributes of the decision problem. We say that a utility function u is in the class $\mathcal{U}^{\mathcal{G}}$ of utilities* **compatible** *to the graph $\mathcal{G}$, if:*

- *the ancestral components of the subgraph $\mathcal{G}'$ of $\mathcal{G}$ induced by $\{\boldsymbol{r}_i : i \in U\}$ are GAI;*

- *the independence structure associated to an ancestral component of $\mathcal{G}'$, A_i say, can be described by a subgraph of the utility diagram obtained by the following procedure:*

  1. derive the subgraph $\mathcal{G}''$ of $\mathcal{G}'$ induced by $\{\boldsymbol{r}_j : j \in A_i\}$;

  2. reverse the direction of every edge in $\mathcal{G}''$;

Example 4.1.9. Consider the DAG in Figure 4.1 and let $U = [4]_1$. Recall that $\{4\}$ and $\{2,3\}$ are in different ancestral components and therefore in the class of compatible utility factorisations $\boldsymbol{r}_4$ and $\{\boldsymbol{r}_2, \boldsymbol{r}_3\}$ are GAI. The two utility diagrams derived following the procedure in Definition 4.1.8 would then correspond to the following two networks: one including only the vertex $\boldsymbol{r}_4$, the other with vertices $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ and an edge from $\boldsymbol{r}_3$ to $\boldsymbol{r}_2$.

We next show that compatible utility functions enjoy a useful factorisation. Recall that D_i is the index set of the descendants of $\boldsymbol{r}_i$.

Proposition 4.1.10. *A utility function u over $\boldsymbol{r}_U$ compatible with the graph $\mathcal{G}$ can be written as*

$$u^{\mathcal{G}}(\boldsymbol{r}_U) = \sum_{i\in Le} u_i^{\mathcal{G}}(\boldsymbol{r}_{A_i}) = \sum_{i\in Le} \sum_{\boldsymbol{r}_{A_i}^{*0} \in \mathcal{R}_{A_i}^{*0}} u\left(\boldsymbol{r}_{A_i}^{*0}\right) \prod_{j\in A_i} g_j\left(\boldsymbol{r}_j \mid \boldsymbol{r}_{D_j}^{*0}, \boldsymbol{r}_{\Pi_j}^{*0}\right), \qquad (4.1.3)$$

where

$$g_j\left(\boldsymbol{r}_j \mid \boldsymbol{r}_{D_j}^{*0}, \boldsymbol{r}_{\Pi_j}^{*0}\right) = \begin{cases} u\left(\boldsymbol{r}_j \mid \boldsymbol{r}_{D_j}^{*0}, \boldsymbol{r}_{\Pi_j}^{*0}\right), & \text{if } \boldsymbol{r}_j = \boldsymbol{r}_j^* \text{ in } u\left(\boldsymbol{r}_{A_i}^{*0}\right), \\ \breve{u}\left(\boldsymbol{r}_j \mid \boldsymbol{r}_{D_j}^{*0}, \boldsymbol{r}_{\Pi_j}^{*0}\right), & \text{otherwise.} \end{cases}$$

Proof. This result follows by first using the GAI factorisation in equation (2.4.7) over the ancestral components and then applying the expansion in equation (2.4.10) to each of these ancestral terms in decreasing order over their indices. The utility independence structure associated to each directed utility diagram then guarantees the result follows, as specified by Lemma 2.4.26. □

Each of the utilities $u_i^{\mathcal{G}}$ in equation (4.1.3) is the product of terms $u(\boldsymbol{r}_{A_i}^{*0})$, corresponding to functions of criterion weights [French and Rios Insua, 2000] and $g_j(\boldsymbol{r}_j \mid \boldsymbol{r}_{D_j}^{*0}, \boldsymbol{r}_{\Pi_j}^{*0})$ which, from Lemma 2.4.26, is a function of $\boldsymbol{r}_j$ only since the attributes $\boldsymbol{r}_{D_j}$ and $\boldsymbol{r}_{\Pi_j}$ are fixed to a certain instantiation. An important consequence of Proposition 4.1.10 in this multiexpert setting is that a compatible utility factorisation is a function of the criterion weights and utilities whose arguments are overseen by individual panels only.

Example 4.1.11. Under the conditions of Example 4.1.9, the most general compatible utility factorisation for the DAG in Figure 4.1 can be written as

$$u(\boldsymbol{r}_2, \boldsymbol{r}_3, \boldsymbol{r}_4) = u_4(\boldsymbol{r}_4) + u(\boldsymbol{r}_2^*, \boldsymbol{r}_3^*) u(\boldsymbol{r}_3 \mid \boldsymbol{r}_2^*) u(\boldsymbol{r}_2 \mid \boldsymbol{r}_3^*) + u(\boldsymbol{r}_2^*, \boldsymbol{r}_3^0) \breve{u}(\boldsymbol{r}_3 \mid \boldsymbol{r}_2^*) u(\boldsymbol{r}_2 \mid \boldsymbol{r}_3^0)$$
$$+ u(\boldsymbol{r}_2^0, \boldsymbol{r}_3^*) u(\boldsymbol{r}_3 \mid \boldsymbol{r}_2^0) \breve{u}(\boldsymbol{r}_2 \mid \boldsymbol{r}_3^*) + u(\boldsymbol{r}_2^0, \boldsymbol{r}_3^0) \breve{u}(\boldsymbol{r}_3 \mid \boldsymbol{r}_2^0) \breve{u}(\boldsymbol{r}_2 \mid \boldsymbol{r}_3^0).$$

We now make the following assumption.

Structural Assumption 4.1.12. *The collective is able to identify an agreed compatible multiattribute utility decomposition over $\boldsymbol{r}_U$ within the class $\mathcal{U}^{\mathcal{G}}$ and to elicit the common $u(\boldsymbol{r}_{A_i}^{*0})$, $i \in Le$.*

4.1.1.4 Distributed Expected Utilities. Under the structural assumptions introduced above, which specify the qualitative structure of the decision problem, the expected utility function factorises into separate factors of the beliefs that particular individual panels can provide themselves. To show this, let, for $t \in [T-1]$,

$$\bar{u}^T \left(\boldsymbol{y}^{T-1}, \boldsymbol{d}^T \right) = \int_{\boldsymbol{y}} u^{\mathcal{G}} \left(\boldsymbol{r}_U \right) f \left(\boldsymbol{y}(T) \mid \boldsymbol{y}^{T-1}, \boldsymbol{d}^T \right) \mathrm{d}\boldsymbol{y}(T),$$
$$\bar{u}^{t-1} \left(\boldsymbol{y}^{t-1}, \boldsymbol{d}^T \right) = \int_{\boldsymbol{y}} \bar{u}^t \left(\boldsymbol{y}^t, \boldsymbol{d}^T \right) f \left(\boldsymbol{y}(t) \mid \boldsymbol{y}^{t-1}, \boldsymbol{d}^T \right) \mathrm{d}\boldsymbol{y}(t).$$

These two terms correspond to the expected utility scores after marginalisation steps have been performed over all the variables with time index bigger or equal than t in the algorithms we define below.

We now show that any function $\bar{u}^t$, $t \in [T]$, can be deduced recursively as a function of the individual panels' statements. Recall that S_j is the set of the indices of the sons of $\boldsymbol{Y}_j(t)$ in a DAG $\mathcal{G}$ (see Appendix B).

Theorem 4.1.13. *Under Structural Assumptions 4.1.3, 4.1.4, 4.1.6 and 4.1.12, $\bar{u}^t$, for $t \in [T]$, can be written as*

$$\bar{u}^t = \int_{\boldsymbol{\mathcal{Y}}_1} \tilde{u}_{t,1} \left(\boldsymbol{y}_1^t, \boldsymbol{d}^T \right) f_{t,1} \left(\boldsymbol{y}_1(t) \mid \boldsymbol{y}_1^{t-1}, \boldsymbol{d}_1^{t-1} \right) \mathrm{d}\boldsymbol{y}_1(t), \qquad (4.1.4)$$

where

$$
\tilde{u}_{t,i}\left(\boldsymbol{y}_{A_i}^t, \boldsymbol{d}^T\right) =
\begin{cases}
\sum_{\boldsymbol{r}_{A_i}^{*0} \in \boldsymbol{\mathcal{R}}_{A_i}^{*0}} u\left(\boldsymbol{r}_{A_i}^{*0}\right) \prod_{j \in A_i} g_j\left(\boldsymbol{r}_j \mid \boldsymbol{r}_{D_j}^{*0}, \boldsymbol{r}_{\Pi_j}^{*0}\right), & i \in Le,\ t = T, \\[2em]
\hat{u}_{t+1,i}\left(\boldsymbol{y}_{A_i}^t, \boldsymbol{d}^T\right), & i \in Le,\ t \neq T, \\[2em]
\sum_{j \in S_i} \bar{u}_{t,j}\left(\boldsymbol{y}_{A_j'}^t, \boldsymbol{y}_j^{t-1}, \boldsymbol{d}^T\right), & otherwise,
\end{cases}
$$

$$(4.1.5)$$

$$
\bar{u}_{t,i}\left(\boldsymbol{y}_{A_j'}^t, \boldsymbol{y}_j^{t-1}, \boldsymbol{d}^T\right) = \int_{\boldsymbol{\mathcal{Y}}_i} \tilde{u}_{t,i}\left(\boldsymbol{y}_{A_i}^t, \boldsymbol{d}^T\right) f_{t,i}\left(\boldsymbol{y}_i(t) \mid \boldsymbol{d}_{A_i'}^t, \boldsymbol{y}_{\Pi_i}^t, \boldsymbol{y}_i^{t-1}\right) \mathrm{d}\boldsymbol{y}_i(t), \qquad (4.1.6)
$$

and $\hat{u}_{t,i}$ is uniquely defined as the function for which

$$
\bar{u}^{t+1} = \sum_{i \in Le} \hat{u}_{t,i}(\boldsymbol{y}_{A_i}^t, \boldsymbol{d}^T), \qquad (4.1.7)
$$

The proof of this theorem is provided in Appendix A.3.1.

Theorem 4.1.13 assures that the expected utility scores for any policy of an IDSS respecting our new structural assumptions can be computed. We proceed to show in Section 4.4 below that the computation of these scores can be performed in a distributed fashion.

We note here that again the actual algebraic form of the terms in equations (4.1.4)-(4.1.7) is not fundamental to the construction of a coherent distributed IDSS. This form depends on the individual panels' agreements concerning the quantities under their particular jurisdiction. Importantly, however, any $\bar{u}^t$ can be written as a function of these terms, whatever they are. Its computation, as we show in Section 4.3, can therefore be obtained through a propagation algorithm, guided by the topology of the DAG of the DDM, between each individual panel and the SB.

The quantities appearing in Theorem 4.1.13 are fundamental to later developments of this chapter. So we now discuss their interpretation. The definition of $\tilde{u}_{t,i}$ in equation (4.1.5) depends on whether or not $\boldsymbol{Y}_i(t)$ is a leaf vertex of the time slice DAG $\mathcal{G}$. In the former case, for $t = T$, this corresponds to the utility function over the appropriate ancestral component, whilst, if $t \neq T$, this is simply equal to $\hat{u}_{t+1,i}$. If $i \notin Le$ then equation (4.1.5) consists of the sum of the terms $\bar{u}_{t,j}$ for $j \in S_i$. Equation (4.1.6) defines $\bar{u}_{t,i}$ which consists of the result of a marginalisation of $\tilde{u}_{t,i}$ with respect to the conditional density function $f_{t,i}$. Finally, the theorem asserts that $\bar{u}^{t+1}$ can be uniquely written as a linear combination of the functions $\hat{u}_{t,i}$, for $i \in Le$. Throughout this chapter, for ease of notation, we use the convention of writing the arguments of the utility functions u and

g in terms of the attributes $\boldsymbol{r}$, whilst for the other functions, e.g. $\tilde{u}_{t,i}$, the arguments are written in terms of the random variables $\boldsymbol{Y}^T$ and the decisions $\boldsymbol{d}^T$.

4.1.2 Components Agreement

It is often recommended that the evaluation of both the conditional utilities and the conditional probabilities should be delegated to groups of individuals best able to compare the efficacy and the likelihood of different value of that attribute [see for example Von Neumann and Morgenstern, 1947, Von Winterfeldt and Edwards, 1986]. We therefore assume the following.

Structural Assumption 4.1.14. *Every expert within a panel G_i agrees on a probabilistic model for the associated component DSS, $f_{t,i}$, $i \in [n]$, $t \in [T]$, as a function of its inputs. In addition every expert in G_i shares a marginal utility function over $\boldsymbol{r}_i$, if $i \in U$.*

It is possible to encourage the experts within a panel to come to these agreements in a variety of ways, appropriate depending on the context, for example through decision conferencing or by following a Delphi Protocol (see Section 2.6.1). Similarly, the probabilistic individual agreement might consist of following certain pooling axioms (see Section 2.6.2.1) or by using agreed software on expert inputs, for example a probabilistic emulator (see Section 2.3.5).

4.2 Relation to the Axioms

In the previous section we followed the construction of an IDSS given in Leonelli and Smith [2015]. This is because the structural assumptions of Section 4.1 more intuitively describe the features of the types of IDSSs we consider in this chapter than the common knowledge axioms of Chapter 3. However, we can simply note here that these structural assumptions represent an instance of the axioms of Chapter 3.

The utility consensus in Axiom 3.2.2 simply corresponds in this chapter to the agreement of a utility factorisation within the class of compatible utility functions as specified in Structural Assumption 4.1.12.

The policy consensus in Axiom 3.2.1 consists of Structural Assumption 4.1.4 defining a multivariate and asymmetric decision space.

The structural consensus in Axiom 3.2.3 is specified by Structural Assumptions 4.1.3 and 4.1.6, which respectively consist of the agreement of using a DDM model and of a shared set of extended conditional independences.

Lastly, the quantitative delegation consensus in Axiom 3.2.5 coincides with Structural Assumption 4.1.14 where in this case panels agree to deliver conditional density functions of the agreed DDM and conditional utilities of the shared compatible utility function.

4.3   Distributed Algorithms

Now that the structure of the IDSS has been fully defined, we can proceed to discuss the computation of the expected utilities through propagation over the network of ESs. We first introduce an algorithm which includes partial optimisation steps to deduce an optimal expected utility score. We then consider two special cases of this algorithm. The first one does not include optimisation steps and computes the expected utility score of a specific policy, whilst the second works over a non-dynamic network of ESs.

4.3.1   Dynamic Optimisation Algorithm

In contrast to the quantities defined in equations (4.1.4)-(4.1.7), which compute the expected utility score of a *particular policy*, we now include additional optimisation steps to the algorithms. These enable us to identify an optimal policy. In fact we can use exactly the same propagation procedure in this case. For this slight generalisation we need to first define a new quantity, $u_{t,i}^*$, which accounts for optimisations over decision spaces. Let

$$u_{t,i}^* \left(\boldsymbol{y}_{A_i}^t, \boldsymbol{d}^{t-1}, \boldsymbol{d}_{A_i'}(t) \right) = \max_{\mathcal{D}_i(t)} \tilde{u}_{t,i} \left(\boldsymbol{y}_{A_i}^t, \boldsymbol{d}^{t-1}, \boldsymbol{d}_{A_i}(t) \right). \qquad (4.3.1)$$

This function is an optimised version, over the decision space $\mathcal{D}_i(t)$, of $\tilde{u}_{t,i}$. We also let $\bar{u}_{t,i}^*$ be the result of the marginalisation of $u_{t,i}^*$. Specifically,

$$\bar{u}_{t,i}^* \left(\boldsymbol{y}_{A_i'}^t, \boldsymbol{y}_i^{t-1}, \boldsymbol{d}^{t-1}, \boldsymbol{d}_{A_i'}(t) \right) = \int_{\boldsymbol{\mathcal{Y}}_i} u_{t,i}^* \left(\boldsymbol{y}_{A_i}^t, \boldsymbol{d}^{t-1}, \boldsymbol{d}_{A_i'}(t) \right) f_{t,i} \mathrm{d}\boldsymbol{y}_i(t). \qquad (4.3.2)$$

Before illustrating the algorithm using the network of Figure 4.1, we introduce a new notation which is also used in the formal algorithms below. We let G_i: or SB: denote the entity that is responsible for the corresponding operation, whilst we represent with $\longrightarrow G_i$ or $\longrightarrow$ SB the fact that panel G_i and the SB, respectively, receives the value of an appropriate function. So, for instance, G_i: $\tilde{u}_{t,i} \longrightarrow$ SB denotes that panel G_i computes the function $\tilde{u}_{t,i}$ and communicates its value to the SB.
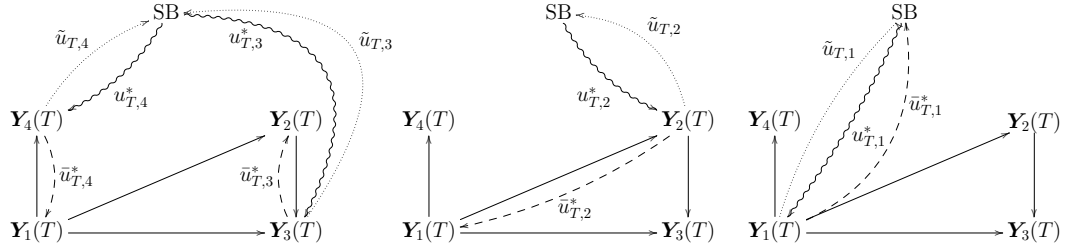
FIGURE 4.3: Collective optimal expected utility algorithm over the last time slice of the network in Figure 4.1.
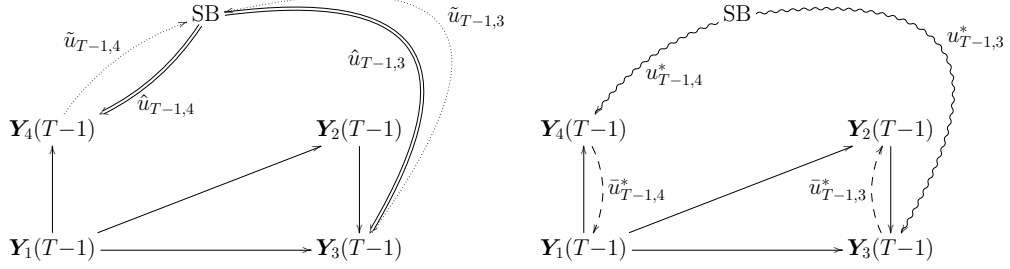
FIGURE 4.4: Collective optimal expected utility algorithm over the $T-1$ time slice of the network in Figure 4.1.

The following example describes the steps of the distributed propagation algorithm for IDSSs whose structural consensus includes a DDM with associated DAG in Figure 4.1.

Example 4.3.1. The algorithm starts from the leaves of the last (time T) time slice DAG and assumes that each panel overseeing a leaf of the time slice DAG, $\boldsymbol{Y}_i(T)$ say, has been provided with the term $u_i^{\mathcal{G}}$. Panels G_i: $\tilde{u}_{T,i} \longrightarrow$ SB, for $i \in [4]_2$. Note that in this case $\tilde{u}_{T,i}$ simply corresponds to $u_i^{\mathcal{G}}$. This step is represented by the dotted arrows on the left network of Figure 4.3 from $\boldsymbol{Y}_3(T)$ and $\boldsymbol{Y}_4(T)$ to the SB. Then SB: $u_{T,i}^* \longrightarrow G_i$ as in equation (4.3.1) for $i \in [4]_2$. This is depicted by the curly arrows in the left network of Figure 4.3. At this stage G_4: $\bar{u}_{T,4}^* \longrightarrow G_1$ and G_3: $\bar{u}_{T,3}^* \longrightarrow G_2$, since $\boldsymbol{Y}_2(T)$ is the father of $\boldsymbol{Y}_3(T)$ and $\boldsymbol{Y}_1(T)$ is the father of $\boldsymbol{Y}_4(T)$ (recall that F_i is the index of the father of $\boldsymbol{Y}_i(T)$). These two operations are described by the dashed arrows on the left network of Figure 4.3.

Now G_2: $\tilde{u}_{T,2} \longrightarrow$ SB, where $\tilde{u}_{T,2} = \bar{u}_{T,3}^*$ since $\boldsymbol{Y}_2(T)$ has only one son. Then, as before, SB: $u_{T,2}^* \longrightarrow G_2$ and G_2: $\bar{u}_{T,2}^* \longrightarrow G_1$, since $\boldsymbol{Y}_1(T)$ is the father of $\boldsymbol{Y}_2(T)$. The whole process is depicted by the network in the middle of Figure 4.3, where, as before, a dotted arrow is associated to $\tilde{u}_{T,i}$, a curly arrow to $u_{T,i}^*$ and a dashed one to $\bar{u}_{T,i}^*$.

Because $\boldsymbol{Y}_1(T)$ is the father of both $\boldsymbol{Y}_2(T)$ and $\boldsymbol{Y}_4(T)$, now G_1: $\tilde{u}_{T,1} \longrightarrow$ SB, by simply adding $\bar{u}_{T,2}^*$ and $\bar{u}_{T,4}^*$, received from panels G_2 and G_4 respectively. Panel G_1 then repeats the same procedure as the other panels, with the only difference that $\bar{u}_{T,1}^* \longrightarrow$ SB and not to another panel, since G_1 oversees the unique root of the DAG. This is depicted by the dashed arrow in the right network of Figure 4.3.

Algorithm 4.3.1: COLLECTIVE OPTIMAL EXPECTED UTILITY$(\boldsymbol{u}, \boldsymbol{f}, \mathcal{G})$

$$
\mathbf{do} \begin{cases} \mathbf{for}\ t \leftarrow T\ \mathbf{downto}\ 1 & (1) \\ \mathbf{do} \begin{cases} \mathbf{for}\ i \leftarrow n\ \mathbf{downto}\ 1 & (2) \\ \begin{cases} \mathbf{if}\ i \in Le & (3) \\ \begin{cases} \mathbf{if}\ t = T & (4) \\ \left\{ G_i : \tilde{u}_{t,i} = u_i^{\mathcal{G}} \longrightarrow SB \right. & (5) \\ \quad \mathbf{else}\ G_i : \tilde{u}_{t,i} = \hat{u}_{t,i} \longrightarrow SB & (6) \end{cases} \\ \quad \mathbf{else}\ G_i : \tilde{u}_{t,i} = \sum_{j \in S_i} \bar{u}_{t,j}^* \longrightarrow SB & (7) \\ SB : u_{t,i}^* = \max_{\mathcal{D}_i(t)} \tilde{u}_{t,i} \longrightarrow G_i & (8) \\ \mathbf{if}\ (i \neq 1) & (9) \\ \quad \mathbf{then}\ \left\{ G_i : \bar{u}_{t,i}^* = \int_{\boldsymbol{\mathcal{Y}}_i} u_{t,i}^* f_{t,i} \mathrm{d}\boldsymbol{y}_i(t) \longrightarrow G_{F_i} \right. & (10) \\ \quad \mathbf{else} \begin{cases} G_i : \bar{u}_{t,i}^* = \int_{\boldsymbol{\mathcal{Y}}_i} u_{t,i}^* f_{t,i} \mathrm{d}\boldsymbol{y}_i(t) \longrightarrow SB & (11) \\ \mathbf{if}\ t \neq 1 & (12) \\ \quad \mathbf{then} \begin{cases} \mathbf{for\ each}\ j \in Le & (13) \\ \quad \mathbf{do}\ \left\{ SB : \hat{u}_{t-1,j} \longrightarrow G_j \right. & (14) \end{cases} \\ \quad \mathbf{else}\ SB : u_0^* = \max_{\mathcal{D}(0)} \bar{u}_{t,i}^* & (15) \end{cases} \end{cases} \end{cases} \end{cases}
$$

Theorem 4.1.13 states that $\bar{u}_{T,1}^* = \bar{u}^T$ is equal to the sum of the terms $\hat{u}_{T-1,i}$, $i \in Le$. So, if i is the index of a leaf vertex, SB: $\hat{u}_{T-1,i} \longrightarrow G_i$. This is denoted in the left network of Figure 4.4 by the double arrows. Panels G_i: $\tilde{u}_{T-1,i} \longrightarrow$ SB, $i \in Le$, where $\tilde{u}_{T-1,i} = \hat{u}_{T-1,i}$. From this stage on, the message passing algorithm copies the calculations and the actions of the previous time slice. So the arrows in Figure 4.4 match the ones on the left network of Figure 4.3. The algorithm repeats the same sequence depicted by the dashed, curly, dotted and double arrows in Figure 4.3 and 4.4, until it reaches the root vertex of the first time slice. When this happens the SB, after receiving $\bar{u}_{1,1}^*$ from panel G_1, computes a final optimisation step over the decision space $\mathcal{D}(0)$. The algorithm has now been completed and can return the expected utility score of the optimal sequence of decisions.

Having described the algorithm on the running example, we now introduce it for a generic DDM and in particular for more realistic scenarios. Specifically, this algorithm takes as inputs the utilities $u_i^{\mathcal{G}}$ associated to the ancestral components of $\mathcal{G}$, denoted as $\boldsymbol{u}$, all the conditional density functions $f_{t,i}$, denoted as $\boldsymbol{f}$, and all the information concerning the DAG $\mathcal{G}$. A formal definition of the algorithm can be found in Algorithm 4.3.1, which is henceforth called **collective optimal expected utility algorithm**. For simplicity, we have left implicit the arguments of various quantities the panels and the SB communicate to each other.

Theorem 4.3.2. *Under Structural Assumptions 4.1.3, 4.1.4, 4.1.6, 4.1.12 and 4.1.14, Algorithm 4.3.1 produces an optimal expected utility score resulting from a unique Bayesian probability model, informed only by the individual judgements delivered by the panels.*

The proof of this theorem is provided in Appendix A.3.2.

We highlight the relations between our algorithm and standard backward inductive evaluation in Section 4.3.3 below.

The utility function is often a polynomial function in its attributes. When this is so, its expectation is a polynomial in which the indeterminates are, in the continuous case, low order moments. This can dramatically simplify the message passing algorithm for computing the optimal policy as we illustrate below. As a result the IDSS often needs as inputs only a few low order moments to work coherently. This is the main topic of Chapter 5. Even in rather complex domains this in turn means that we can expect the algorithms defined above to be almost instantaneous if each component module can produce the values of these uncertainties under various policy choices efficiently. A study of the polynomial structure of expected utilities in the rather more complex discrete domain is reviewed in Section 5.6.1 and in Leonelli et al. [2015a].

4.3.2 The Expected Utility of a Policy

Algorithm 4.3.1 provides an operational guideline on how to compute the score associated to an optimal policy. Recall however that the aim of a DSS is not only to identify the decisions with highest expected utilities, but also to provide explanations and the reasoning behind the outputs it provides [French et al., 2009]. It is therefore also relevant to compute the expected utility score associated with any policy that might be adopted. These scores then allow decision centres to compare the different available options in more detail, possibly following the route described in Figure 3.2. To accommodate this feature a simple variant of Algorithm 4.3.1 not including any optimisation steps is presented in Algorithm 4.3.2, henceforth called **collective expected utility** algorithm. Note that this is derived from Algorithm 4.3.1 by dropping lines (8) and (15) and by replacing $\bar{u}_{t,j}^*$ with $\bar{u}_{t,j}$.

4.3.3 The non Dynamic Case

In some domains it can be more appropriate to model decision problems using a non-dynamic probabilistic model, such as a BN. Within the IDSS framework, this is possible

---

Algorithm 4.3.2: COLLECTIVE EXPECTED UTILITY$(\boldsymbol{u}, \boldsymbol{f}, \mathcal{G}, \boldsymbol{d})$

---

for $t \leftarrow T$ **downto** 1

$\quad$**do** $\begin{cases} \textbf{for } i \leftarrow n \textbf{ downto } 1 \\ \quad \textbf{do} \begin{cases} \textbf{if } i \in Le \\ \quad \begin{cases} \textbf{if } t = T \\ \quad \left\{ G_i : \tilde{u}_{t,i} = u_i^{\mathcal{G}} \longrightarrow SB \right. \\ \quad\quad \textbf{else } G_i : \tilde{u}_{t,i} = \hat{u}_{t,i} \longrightarrow SB \\ \quad \textbf{else } G_i : \tilde{u}_{t,i} = \sum_{j \in S_i} \bar{u}_{t,j} \longrightarrow SB \end{cases} \\ \textbf{if } (i \neq 1) \\ \quad \textbf{then } \left\{ G_i : \bar{u}_{t,i} = \int_{\boldsymbol{\mathcal{Y}}_i} \tilde{u}_{t,i} f_{t,i} \mathrm{d}\boldsymbol{y}_i(t) \longrightarrow G_{F_i} \right. \\ \quad \textbf{else } \begin{cases} G_i : \bar{u}_{t,i} = \int_{\boldsymbol{\mathcal{Y}}_i} \tilde{u}_{t,i} f_{t,i} \mathrm{d}\boldsymbol{y}_i(t) \longrightarrow SB \\ \textbf{if } t \neq 1 \\ \quad \textbf{then } \begin{cases} \textbf{for each } j \in Le \\ \quad \textbf{do } \left\{ SB : \hat{u}_{t-1,j} \longrightarrow G_j \right. \end{cases} \end{cases} \end{cases} \end{cases}$

---

by simply adapting Algorithm 4.3.1 to the non-dynamic case. Algorithm 4.3.3, which we call henceforth **non-dynamic optimal EU** algorithm, shows how this can be done. Note that we adapt the notation to the non-dynamic case by dropping the dependence on the time-varying index: therefore we are now using the standard notation of non-dynamic domains of the previous chapters. Although some symbols in the algorithm are new, these are self-explanatory and follow straightforwardly from the dynamic one. It is easy to notice that this algorithm works, since the last time slice of a dynamic DDM alone can be thought of as a non dynamic network. For this time slice, Theorem 4.3.2 guarantees the algorithm computes exact expected utility scores.

In the non dynamic case it is easier to highlight the relationships between our algorithms and standard backward induction evaluation of IDs introduced in Proposition 2.5.10. Consider the time slice DAG of Figure 4.2, which can be related to a non-dynamic problem. The conditions imposed our structural assumptions guarantee that a standard evaluation can be performed over each ancestral set of the DAG considering only the utility function over that component. Furthermore, as stated by Theorem 4.1.13, the evaluation of each of these ancestral components can be distributed to the different panels since the expected utility factorises accordingly.

Algorithm 4.3.3: NON-DYNAMIC COLLECTIVE OPTIMAL $\mathrm{EU}(\boldsymbol{u}, \boldsymbol{f}, \mathcal{G})$

---

for $i \leftarrow n$ **downto** 1

$$
\mathbf{do}\begin{cases}
\mathbf{if}\ i \in Le \\
\quad \left\{ G_i : \tilde{u}_i = u_i^{\mathcal{G}} \longrightarrow SB \right. \\
\quad \mathbf{else}\ G_i : \tilde{u}_i = \sum_{j \in S_i} \bar{u}_j^* \longrightarrow SB \\
SB : u_i^* = \max_{\mathcal{D}_i} \tilde{u}_i \longrightarrow G_i \\
\mathbf{if}\ (i \neq 1) \\
\quad \mathbf{then}\ \left\{ G_i : \bar{u}_i^* = \int_{\boldsymbol{y}_i} u_i^* f_i \mathrm{d}\boldsymbol{y}_i \longrightarrow G_{F_i} \right. \\
\quad \mathbf{else}\ \begin{cases} G_i : \bar{u}_i^* = \int_{\boldsymbol{y}_i} u_i^* f_i \mathrm{d}\boldsymbol{y}_i \longrightarrow SB \\ SB : u_0^* = \max_{\mathcal{D}_0} \bar{u}_i^* \end{cases}
\end{cases}
$$

4.4 Examples

The previous section formally presented distributed propagation algorithms for IDSSs. These work in general and for any network of ESs that respect the structural assumptions of Section 4.1. We now illustrate how the different panels of experts can communicate with each other through a suite of different IDSSs concerning the policies after an accidental release of contaminants at a nuclear power plant.

4.4.1 A Simple Bayesian Network

Consider a simple example consisting of two dependent random variables, Y_1 and Y_2, overseen by two different panels. The variable Y_1 measures the number of people experiencing negative symptoms after a nuclear accident, whilst Y_2 estimates the financial implications of that accident. Assume both variables to be continuous and defined through the simple DAG linear regressions in equation (2.3.4). Therefore $Y_1 = \theta_{01} + \varepsilon_1$ and $Y_2 = \theta_{02} + \theta_{12} Y_1 + \varepsilon_2$, where ε_i, $i \in [2]$, is a random error with mean zero and variance ψ_i. Further let $\mathbb{E}(\theta_{ij} \mid d) = a_{ij}$, $\mathbb{V}(\theta_{01} \mid d) = c_{01}$ and $E(\psi_1 \mid d) = b_1$, where d denotes an available policy. This is a first example of a partially defined IDSS that we study in depth in Chapter 5. This situation can be described by the simple BN in Figure 4.5 where in this case we have also added the parameters and the hyperparameters as vertices [this is often done in practice, as discussed in O'Hagan and Forster, 2004]. Note that the edge set of this DAG implies panel independence a priori since there are no edges between $\{\theta_{01}, \psi_1\}$ and $\{\theta_{02}, \theta_{12}, \psi_2\}$. Furthermore this DAG implies local independence since there are no edges between any two parameters.
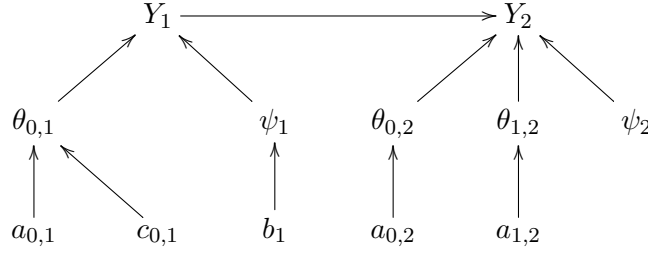
FIGURE 4.5: Bayesian network describing the example in Section 4.4.1, including the parameters and the hyperparameters in its vertex set.

Assume $r_i = y_i$, $i \in [2]$, and $U(r_1, r_2) = y_1 y_2$. Suppose also the decision space simply consists of an initial decision space $\mathcal{D}(0) = \{d_1, d_2\}$, where d_1 denotes evacuation, whilst d_2 corresponds to the delivery of protective measures. For ease of notation, suppress the dependence on decision $d \in \mathcal{D}(0)$.

Since in this setting the DAG has a unique ancestral component, we have that $\tilde{u}_2 = y_1 y_2$ and in order to compute $\bar{u}_2$, panel G_2 simply needs to compute the expected value of this function. By sequentially applying the tower rule for expectations and the properties of the expectation operator in Propositions 2.1.5 and 2.1.6 respectively, and letting $\boldsymbol{\theta}$ denote the overall parameter vector, panel G_2 computes

$$\mathbb{E}(Y_1 Y_2) = \mathbb{E}(\mathbb{E}(Y_1 Y_2 \mid \boldsymbol{\theta}, Y_1)) = \mathbb{E}((\theta_{02} + \theta_{12} Y_1) Y_1) = \mathbb{E}(\theta_{02} Y_1 + \theta_{12} Y_1^2)$$
$$= \mathbb{E}(\theta_{02})\mathbb{E}(Y_1) + \mathbb{E}(\theta_{12})\mathbb{E}(Y_1^2) = \bar{u}_2$$

The function $\bar{u}_2$ coincides with $\tilde{u}_1$ since Y_2 is the only son of Y_1. Therefore now panel G_1 can compute $\bar{u}_1$ as

$$\mathbb{E}(\tilde{u}_1) = \mathbb{E}(\theta_{02})\mathbb{E}(Y_1) + \mathbb{E}(\theta_{12})\mathbb{E}(Y_1^2) = \mathbb{E}(\theta_{02})\mathbb{E}(\mathbb{E}(Y_1 \mid \boldsymbol{\theta})) + \mathbb{E}(\theta_{12})(\mathbb{E}(Y_1)^2 + \mathbb{V}(Y_1))$$
$$= \mathbb{E}(\theta_{02})\mathbb{E}(\theta_{01}) + \mathbb{E}(\theta_{12})(\mathbb{E}(\mathbb{E}(Y_1 \mid \boldsymbol{\theta}))^2 + \mathbb{V}(\mathbb{E}(Y_1 \mid \boldsymbol{\theta})) + \mathbb{E}(\mathbb{V}(Y_1 \mid \boldsymbol{\theta})))$$
$$= \mathbb{E}(\theta_{02})\mathbb{E}(\theta_{01}) + \mathbb{E}(\theta_{12})(\mathbb{E}(\theta_{01})^2 + \mathbb{V}(\theta_{01}) + \mathbb{E}(\psi_1))$$
$$= a_{02} a_{01} + a_{12}(a_{01}^2 + c_{01} + b_1) \tag{4.4.1}$$

The form of the expected utility in equation (4.4.1) informs the IDSS about the beliefs panels need to deliver. Suppose panel G_1 delivers the beliefs $\Psi_1^{\boldsymbol{y}}(d_1) = \Psi_1^{\boldsymbol{\theta}}(d_1) = \{a_{01} = 2, c_{01} = 4, b_1 = 4\}$ and $\Psi_1^{\boldsymbol{y}}(d_2) = \Psi_1^{\boldsymbol{\theta}}(d_2) = \{a_{01} = 2.5, c_{01} = 1, b_1 = 2\}$, whilst panel G_2 delivers $\Psi_2^{\boldsymbol{y}}(d_1) = \Psi_2^{\boldsymbol{\theta}}(d_1) = \{a_{02} = 2.5, a_{12} = 2\}$ and $\Psi_2^{\boldsymbol{y}}(d_2) = \Psi_2^{\boldsymbol{\theta}}(d_2) = \{a_{02} = 2, a_{12} = 2\}$.

A short note on the values assigned to this hyperparameters. The hyperparameter a_{01} corresponds to the expectation of the expected number of people with adverse symptoms, whilst a_{02} represents the expectation of expected amount of financial cost. It is then

reasonable to provide a higher value for a_{01} in the case of protective measures than in the evacuation scenario. By the same logic, even though protective measures may lead to more expenses for medical cures, the costs of an evacuation are incredibly high and significantly greater than for the other available option. The hyperparameter a_{12} represents the strength of the causal relationship between the two variables. A priori, G_2 may not be very well informed about this quantity and decide to provide the same value for the two decisions available. The expectation of the variance of the number of people with negative symptoms is b_1: in the case of delivering protective measures this value is definitively lower, since it is easier to provide shelter to most of the population. On the other hand during an evacuation, it may not be possible to help all the people in danger and only a subset of them may be moved to safer areas: as a consequence, the expectation of the variance of Y_1 for decision d_1 is higher. The last hyperparameter c_{01} corresponds to the variance of the expectation of Y_1 and, by the same argument, it is higher in the evacuation case than in the protective measures one.

Note that the IDSS would have not been able to compute the expected utility of equation (4.4.1) if panels had only delivered the mean estimates of the variables under their jurisdiction. The quantities c_{01} and b_1 represent levels of uncertainty concerning these means. If we just plug in the expectations of relevant means and set equal to zero the above measures of uncertainty, we obtain a spurious evaluation of an expected utility score corresponding to

$$a_{01}a_{02} + a_{12}a_{01}^2.$$

In this case, plugging in the values provided by the panels, the expected utility for the evacuation scenario is 13, whilst the expected value for delivering protective measures is 17.5: the optimal choice, when no variation in the estimates is allowed, is to evacuate the population.[1] Consider now the actual expected utility expression in equation (4.4.1), including the second-order uncertainty. In this situation the expected utility of d_1 is 29, whilst the expected utility under d_2 is 23.5, from which it follows that the optimal decision consists of delivering protective measures. The misjudgement of uncertainty in this case has led to an indefensible decision, highlighting the need for a complete uncertainty handling.

4.4.2   A Multiregression Dynamic Model

We now consider a very simple initial example of a dynamic framework, consisting of a dynamic extension of the model of the previous section. Assume a decision centre needs to decide whether to evacuate or not the population of a small village close to

---

[1]Note that in this case the expected utility needs to be minimised since the attributes of the example represent negative consequences.

a nuclear reactor which is contaminating the surrounding area. The course of action can be decided at two time points only with associated binary decision spaces $\mathcal{D}(0)$ and $\mathcal{D}(1)$, where $\mathcal{D}(t) = \{d_1(t), d_2(t)\}$, $t \in [1]^0$. Let $d_1(t)$ be the option of evacuation, whilst $d_2(t)$ corresponds to do nothing, $t \in [1]^0$. Suppose the system is asked to identify the expected utility scores of a specific policy (i.e. no optimisation steps are required) and suppress this dependence.

Assume also that the centre plans to observe the value of two continuous univariate time series $\{Y_1(t)\}_{t \in [2]}$ and $\{Y_2(t)\}_{t \in [2]}$ only at time $t = 1$, i.e. after having committed to a decision in $\mathcal{D}(0)$ but before having to choose an option in $\mathcal{D}(1)$. Assume further that it is only considered relevant for the analysis the next value of these two series at time $t = 2$, which is not observed. As before, at time t, $Y_1(t)$ measures the number of people experiencing negative symptoms after a nuclear accident, whilst $Y_t(2)$ estimates the financial costs of the accident.

Assume that during decision conferences the structural consensus of the IDSS has been agreed by the collective to include an LMDM where $\boldsymbol{Y}_1^2$ is the parent of $\boldsymbol{Y}_2^2$. Further assume the collective agrees that the two series are the only attributes of the problem and that the utility consensus includes a simple linear utility, which can be written as

$$u(y_1(1), y_1(2), y_2(1), y_2(2)) = u(y_1(1)) + u(y_1(2)) + u(y_2(1)) + u(y_2(2)). \qquad (4.4.2)$$

Assume also that the LMDM is specifically defined by the following equations:

- $Y_2(2) = \theta_{12}(2)Y_1(2) + v_2(2);$
- $\theta_{12}(2) = \theta_{12}(1) + w_2(2);$
- $Y_2(1) = \theta_{12}(1)Y_1(1) + v_2(1);$

- $Y_1(2) = \theta_{01}(2) + v_1(2);$
- $\theta_{01}(2) = \theta_{01}(1) + w_1(2);$
- $Y_1(1) = \theta_{01}(1) + v_1(1).$

Recall that in an LMDM the errors $v_i(t)$ and $w_i(2)$ are independent of each other with mean zero and variance $V_i(t)$ and $W_i(2)$, respectively, for $i, t \in [2]$. Each panel individually assumed these variances to be unknown but has provided a prior mean estimate $b_i(t)$ for $V_i(t)$ and $r_i(2)$ for $W_i(2)$. Assume further that the panels provided prior information about the parameter vector at time $t = 1$, such that a_{01} and a_{12} are the mean estimates for $\theta_{01}(1)$ and $\theta_{12}(1)$, respectively. The variances of $\theta_{01}(1)$ and $\theta_{12}(1)$ are elicited to be c_{01} and c_{12} respectively. Assume further that each panel individually believes that the marginal utilities under their jurisdiction are quadratic, so that $u(y_t(i)) = -y_i(t)^2$, $t, i \in [2]$.

Note that all the above panels' specifications are delivered separately, as required by the quantitative delegation consensus. Specifically, the form of the first three equations in the list above is decided by panel G_2, whilst the remaining ones are specified by G_1.

Now that the IDSS has been fully defined for this example, we can show how the algorithm works when the overarching structure is the LMDM. As in the previous case, because of the polynomial form of the utility function, the algorithm consists of a sequential use of the tower rules of moments introduced in Proposition 2.1.6. Because the associated DAG consists of one single ancestral component, we have that $\tilde{u}_{2,2}$ coincides with the rhs of equation (4.4.2). Then, letting $\dot{u}_{2,2} = u(y_1(1)) + u(y_1(2)) + u(y_2(1))$, panel G_2 computes a marginalisation by performing the following steps

$$
\begin{aligned}
E(\tilde{u}_{2,2}) =& \dot{u}_{2,2} - \mathbb{E}(Y_2(2))^2 - \mathbb{V}(Y_2(2)) \\
=& \dot{u}_{2,2} - \mathbb{E}(\mathbb{E}(Y_2(2)|\cdot))^2 - \mathbb{V}(\mathbb{E}(Y_2(2)|\cdot)) - \mathbb{E}(\mathbb{V}(Y_2(2)|\cdot)) \\
=& \dot{u}_{2,2} - \mathbb{E}(\theta_{12}(2)Y_1(2))^2 - \mathbb{V}(\theta_{12}(2)Y_1(2)) - \mathbb{E}(V_2(2)) \\
=& \dot{u}_{2,2} - \mathbb{E}(\theta_{12}(2)Y_1(2))^2 - \mathbb{V}(\mathbb{E}(\theta_{12}(2)Y_1(2)|\cdot)) - \mathbb{E}(\mathbb{V}(\theta_{12}(2)Y_1(2)|\cdot)) - b_2(2) \\
=& \dot{u}_{2,2} - \mathbb{E}(\mathbb{E}(\theta_{12}(2)|\cdot))^2 \mathbb{E}(Y_1(2))^2 - \mathbb{V}(Y_1(2)\theta_{12}(1)) - \mathbb{E}(Y_1(2)^2 W_2(2)) - b_2(2) \\
=& \dot{u}_{2,2} - \mathbb{E}(\theta_{12}(1))^2 \mathbb{E}(Y_1(2))^2 - \mathbb{V}(\mathbb{E}(Y_1(2)\theta_{12}(1)|\cdot)) - \mathbb{E}(\mathbb{V}(Y_1(2)\theta_{12}(1)|\cdot)) \\
& - \mathbb{E}(Y_1(2)^2)\mathbb{E}(W_2(2)) - b_2(2) \\
=& \dot{u}_{2,2} - a_{12}^2 \mathbb{E}(Y_1(2))^2 - \mathbb{V}(Y_1(2)a_{12}) - \mathbb{E}(Y_1(2)^2 c_{12}) \\
& - r_2(2)\mathbb{E}(Y_1(2)^2) - b_2(2) \\
=& \dot{u}_{2,2} - (a_{12}^2 + c_{12} + r_2(2))\mathbb{E}(Y_1(2)^2) - b_2(2) = \bar{u}_{2,2}
\end{aligned}
$$

This expression consists of non-random quantities, terms that are overseen by G_1, or elements associated to the previous time slice only. Therefore, $G_2: \bar{u}_{2,2} \longrightarrow G_1$. Note that $\tilde{u}_{2,1}$ coincides with $\bar{u}_{2,2}$ since G_2 oversees the only son of the vertex overseen by G_1. Letting $\dot{u}_{2,1} = u(y_1(1)) + u(y_2(1)) - b_2(2)$ and $g_{2,1} = 1 + a_{12}^2 + c_{12} + r_2(2)$, panel G_1:

$$
\begin{aligned}
\mathbb{E}(\tilde{u}_{2,1}) =& \dot{u}_{2,1} - g_{2,1}(\mathbb{E}(Y_1(2)^2 + \mathbb{V}(Y_1(2))) \\
=& \dot{u}_{2,1} - g_{2,1}(\mathbb{E}(\mathbb{E}(Y_1(2)|\cdot))^2 + \mathbb{V}(\mathbb{E}(Y_1(2)|\cdot)) + \mathbb{E}(\mathbb{V}(Y_1(2)|\cdot))) \\
=& \dot{u}_{2,1} - g_{2,1}(\mathbb{E}(\theta_{01}(2))^2 + \mathbb{V}(\theta_{01}(2)) + \mathbb{E}(V_1(2))) \\
=& \dot{u}_{2,1} - g_{2,1}(\mathbb{E}(\mathbb{E}(\theta_{01}(2)|\cdot))^2 + \mathbb{V}(\mathbb{E}(\theta_{01}(2)|\cdot)) + \mathbb{E}(\mathbb{V}(\theta_{01}(2)|\cdot)) + b_1(2)) \\
=& \dot{u}_{2,1} - g_{2,1}(\mathbb{E}(\theta_{01}(1))^2 + \mathbb{V}(\theta_{01}(1)) + \mathbb{E}(W_1(2)) + b_1(2)) \\
=& \dot{u}_{2,1} - g_{2,1}(a_{01}^2 + c_{01} + r_1(2) + b_1(2)) = \bar{u}_{2,1}
\end{aligned}
$$

All the variables associated to the second time slice have now been marginalised and the algorithm continues with the marginalisation of $Y_2(1)$. Letting $\dot{u}_{1,2} = u(y_1(1)) - g_{2,1}(a_{01}^2 + c_{01} + r_1(2) + b_1(2)) - b_2(2)$ we have that

$$
\begin{aligned}
\mathbb{E}(\tilde{u}_{1,2}) &= \dot{u}_{1,2} - E(\mathbb{E}(Y_2(1)|\cdot))^2 - \mathbb{V}(\mathbb{E}(Y_2(1)|\cdot)) - \mathbb{E}(\mathbb{V}(Y_2(1)|\cdot)) \\
&= \dot{u}_{1,2} - \mathbb{E}(\theta_{12}(1)Y_1(1))^2 - \mathbb{V}(\mathbb{E}(\theta_{12}(1)Y_1(1)|\cdot)) - \mathbb{E}(\mathbb{V}(\theta_{12}(1)Y_1(1)|\cdot)) - b_2(1) \\
&= \dot{u}_{1,2} - a_{12}^2\mathbb{E}(Y_1(1))^2 - a_{12}^2\mathbb{V}(Y_1(1)) - c_{12}\mathbb{E}(Y_1(1)^2) - b_2(1) \\
&= \dot{u}_{1,2} - (a_{12}^2 + c_{12})\mathbb{E}(Y_1(1)^2) - b_2(1) = \bar{u}_{1,2}.
\end{aligned}
$$

Now letting

$$
\dot{u}_{1,1} = -g_{2,1}(a_{01}^2 + c_{01} + r_1(2) + b_1(2)) - b_2(2) - b_2(1)
$$

and $g_{1,1} = 1 + a_{12}^2 + c_{12}$, G_1:

$$
\begin{aligned}
\mathbb{E}(\tilde{u}_{1,1}) &= \dot{u}_{1,1} - g_{1,1}(\mathbb{E}(\mathbb{E}(Y_1(2)|\cdot))^2 + \mathbb{V}(\mathbb{E}(Y_1(2)|\cdot)) + \mathbb{E}(\mathbb{V}(Y_1(2)|\cdot))) \\
&= \dot{u}_{1,1} - g_{1,1}(a_{01}^2 + c_{01} + b_1(2)) = \bar{u}_{1,1}.
\end{aligned}
$$

The algorithm is now completed and the overall expected utility, coinciding with $\bar{u}_{1,1}$, can be written as

$$
\bar{u}_{1,1} = -\bar{u}(1) - \bar{u}(2) - b_2(1) - b_2(2),
$$

where

$$
\begin{aligned}
\bar{u}(1) &= (1 + a_{12}(1)^2 + c_{12}(1))(a_{01}(1)^2 + c_{01}(1) + b_1(2)), \\
\bar{u}(2) &= (1 + a_{12}(1)^2 + c_{12}(1) + r_2(2))(a_{01}(1)^2 + c_{01}(1) + r_1(2) + b_1(2)).
\end{aligned}
$$

Although straightforward, the recursions in the simplest possible dynamic example enabled us to clearly illustrate symbolically the algorithm in practice. In the following sections we first consider a numerical example and then much more complex dynamic models.

4.4.3 A Discrete Bayesian Network

Consider again a non dynamic domain with four random variables Y_i, $i \in [4]$, each overseen by a different panel of experts and represented by the DAG in Figure 2.1 (the non-dynamic variant of Figure 4.1). Suppose in this case that the variables are binary taking values $\mathcal{Y}_i = [1]^0$, $i \in [4]$, and that these have the same meaning as in Example 2.3.8: specifically Y_1 is an indicator of radioactive contamination in the environment, Y_2 is an indicator of radioactive intake in the population, Y_3 is an indicator of adverse

d_0	d_1	d_2	d_3	$k_d u_d(\cdot)$
1	0	0	0	**0.15**
0	1	0	1	**0.2**
0	1	1	1	**0**
0	0	1	1	0.15
0	0	1	0	**0.2**
0	0	0	1	0.25
0	0	0	0	**0.3**

TABLE 4.1: List of available policies and their utility scores.

d_4	$k_{d_4} u_{d_4}(d_4)$		y_4	$k_4 u_4(y_4)$		y_3	$k_3 u_3(y_3)$		y_2	$k_2 u_2(y_2)$
1	0		1	0		1	0		1	0
0	**0.05**		0	0.1		0	0.45		0	0.1

TABLE 4.2: Specification of the other required utilities.

health effects and Y_4 is an indicator of political disruption in the affected area. Suppose the collective agreed on five binary decision spaces $\mathcal{D}_i$, $i \in [4]^0$. The initial decision $\mathcal{D}_0$ has two options: evacuation or do nothing. The decision after observing the amount of radiation, $\mathcal{D}_1$, consists of providing iodine tablets or not. These can be distributed if evacuation was not chosen as initial decision. The space $\mathcal{D}_2$ consists of the option of setting up a medical field support centre in the area. The centre can be built only if the population was not evacuated. The decision space $\mathcal{D}_3$ is the same as $\mathcal{D}_1$ and the former cannot be equal to not distributing iodine tablets if it was decided to do so after observing the contamination of the area. Finally the decision space $\mathcal{D}_4$ consists of two options: whether or not to proceed with a diplomatic action. Assume further that the collective agrees on a set of extended conditional independences so that the overall density of the IDSS can be written as

$$f(\boldsymbol{y} \mid \boldsymbol{d}) = f_4(y_4 \mid y_1, d_0) f_3(y_3 \mid y_2, y_1, d_0) f_2(y_2 \mid y_1, d_0) f_1(y_1).$$

Note that in particular the distributions of Y_2, Y_3 and Y_4 depend on their parents and on $\mathcal{D}_0$ only.

As before, assume $U = \{2, 3, 4\}$ and suppose the agreed utility function can be written as

$$u^{\mathcal{G}} = k_4 u_4(y_4) + k_3 u_3(y_3) + k_2 u_2(y_2) + k_{d_4} u_{d_4}(d_4) + k_d u_d(d_0, d_1, d_2, d_3). \tag{4.4.3}$$

Note that this utility factorisation is in the compatible class for the DAG associated to this example. The first three terms on the rhs of equation (4.4.3) are provided by

y_2	y_1	d_0	$\mathbb{P}(Y_3 = 1|\cdot)$
1	1	1	0.7
1	0	1	0.7
0	1	1	0.5
0	0	1	0.2
1	1	0	0.9
1	0	0	0.8
0	1	0	0.6
0	0	0	0.3

d_0	y_1	$\mathbb{P}(Y_4 = 1|\cdot)$	$\mathbb{P}(Y_2 = 1|\cdot)$
1	1	0.85	0.7
1	0	0.7	0.9
0	1	0.5	0.1
0	0	0.2	0.2

y_1	$\mathbb{P}(Y_1 = y_1)$
1	0.8
0	0.2

TABLE 4.3: Probability specifications delivered by the panels.

d_0	y_1	$\bar{u}_4(d_0, y_1)$
1	1	0.065
1	0	0.08
0	1	0.1
0	0	0.13

TABLE 4.4: Values of $\bar{u}_4^*$ after the marginalisation step of G_4.

the corresponding panels, whilst we envisage the remaining ones to be agreed by the collective.

Suppose panels and the collective delivered the utility specifications summarised in Tables 4.1 and 4.2. We see that Table 4.1 includes only the policies that respect the constraints introduced above. Further assume that the panels delivered the probabilistic beliefs summarised in Table 4.3.

In this setting the algorithm starts computing $\tilde{u}_3 = u_3 + u_d$ and $\tilde{u}_4 = u_4 + u_{d_4}$. These are sent to the SB that performs the optimisation step over these quantities. The result of these optimisation depends on the available decisions only and is represented in Tables 4.1 and 4.2 by a bold expected utility score. These are the scores of the policies the SB transmits to the appropriate panels in our algorithm. At this stage, following Algorithm 4.3.3, panels G_4 and G_3 perform marginalisation steps to compute respectively

$$\bar{u}_4^* = \sum_{y_4 \in \mathcal{Y}_4} u_4^* \mathbb{P}(Y_4 = y_4 \mid Y_1 = y_1, d_0), \quad \bar{u}_3^* = \sum_{y_3 \in \mathcal{Y}_3} u_3^* \mathbb{P}(Y_3 = y_3 \mid Y_2 = y_2, Y_1 = y_1, d_0).$$

Table 4.4 shows the result of the marginalisation step performed by Panel G_4 which is then communicated to panel G_1. Similarly, Table 4.5 shows $\bar{u}_3^*$ which coincides with $\tilde{u}_2$ since Y_3 is the only son of Y_2 in the associated DAG. Therefore the SB can now identify an optimal policy for each combination of the parent variables and decisions: this again is reported by a bold score on leftmost column of Table 4.5.

d_0	d_1	d_2	d_3	y_1	y_2	$\tilde{u}_2(\cdot)$
1	0	0	0	1	1	**0.465**
0	1	0	1	1	1	**0.605**
0	1	1	1	1	1	0.405
0	0	1	0	1	1	0.605
0	0	0	0	1	1	**0.705**
1	0	0	0	0	1	**0.465**
0	1	0	1	0	1	**0.56**
0	1	1	1	0	1	0.36
0	0	1	0	0	1	0.56
0	0	0	0	0	1	**0.66**
1	0	0	0	1	0	**0.475**
0	1	0	1	1	0	**0.57**
0	1	1	1	1	0	0.37
0	0	1	0	1	0	0.57
0	0	0	0	1	0	**0.67**
1	0	0	0	0	0	**0.34**
0	1	0	1	0	0	**0.435**
0	1	1	1	0	0	0.235
0	0	1	0	0	0	0.435
0	0	0	0	0	0	**0.535**

TABLE 4.5: Value of the functions $\tilde{u}_2$ and $\bar{u}_3^*$ computed respectively by G_2 and G_3.

d_0	d_1	d_2	d_3	d_4	y_1	$\tilde{u}_1(\cdot)$
1	0	0	0	0	1	**0.533**
0	1	0	0	0	1	0.7015
0	0	0	0	0	1	**0.8015**
1	0	0	0	0	0	**0.4325**
0	1	0	0	0	0	0.59
0	0	0	0	0	0	**0.69**

TABLE 4.6: Values of $\tilde{u}_1$ computed by summing $\bar{u}_4^*$ and $\bar{u}_2^*$.

Following Algorithm 4.3.3, panel G_2 performs a marginalisation step and communicates the result to G_1, which can then compute $\tilde{u}_1$ by summing $\bar{u}_2^*$ and $\bar{u}_4^*$ computed respectively by G_2 and G_4. The SB then performs the optimisation step which is summarised in Table 4.6 by the bold scores: there are at this stage four potential optimal policies, one for each combination of y_1 and d_0.

The final two steps of the algorithm consist of a marginalisation over $\mathcal{Y}_1$, performed by panel G_1, and the final optimisation step over the initial decision space $\mathcal{D}_0$, performed by the SB. Given the beliefs delivered by G_1, the IDSS derives the expected utility scores

for the two only remaining policies as:

$$\bar{u}(d_0 = 1, d_1 = 0, d_2 = 0, d_3 = 0, d_4 = 0) = 0.4526,$$
$$\bar{u}(d_0 = 0, d_1 = 0, d_2 = 0, d_3 = 0, d_4 = 0) = 0.7123.$$

Thus, the IDSS, following Algorithm 4.3.3, would suggest that the optimal policy consists of not intervening at any point.

In Section 5.6.1 we briefly demonstrate that evaluations in discrete domains like this one can be performed symbolically and written as a simple arithmetic expression.

4.4.4 Gaussian Recursions

We consider now a dynamic variant of the previous example, where the processes depicted by the DAG in Figure 4.1 have the same meaning as before. As discussed in Chapter 1, many applications we have in mind have a geographical structure, in the sense that many of the values of the required variables are recorded at several locations in an area of interest. This is for example the case in a nuclear emergency, where levels of contamination are collected at many different locations in the surroundings of a power plant. Thus, the processes IDSSs usually deal with are high dimensional. However, the associated utilities are usually low dimensional and can consequently be evaluated transparently. Note that if the impacts of the countermeasures need to be considered at a regional level, it is straightforward to implement these into an IDSS framework. Panels then simply need to provide different scores for the different regions of interest.

Because for real problems the number of equations required to define the problem scales up to an extent where the outworkings of the algorithm are obscured, the example below illustrates how a geographic component can be included into the analysis in the simplest possible case. However in much larger scenarios the calculations are still very feasible and just as straightforward to calculate as in this example because everything is distributed and in closed form. Furthermore, each of the unknown quantities are in practice numbers, rather than algebraic entities, provided by the component DSS and so quick to integrate within the composite system.

For the purpose of this example, suppose the IDSS needs to identify the expected utility score associated to a specific policy and suppress this dependence. Further suppose each vector $\boldsymbol{Y}_i(t) = (Y_i^1(t), \ldots, Y_i^r(t))^{\mathrm{T}}$, $i \in [4]$, $t \in [T]$, is such that $Y_i^l(t)$ is a univariate continuous random variable observed at location $l \in [r]$. The locations are the same for all the time series and do not change through time. To keep this illustration simple we consider here a simple LMDM over a finite time horizon T equal to 2. Specifically, for

$i \in [4]_1$, $t \in [2]$ and $l \in [r]$, we let

$$Y_i^l(t) = \sum_{j \in \Pi_i} \theta_{ji}^l(t) Y_j^l(t) + v_i^l(t), \quad \text{and} \quad \theta_{ji}^l(2) = \theta_{ji}^l(1) + w_{ji}^l(2). \tag{4.4.4}$$

Equation (4.4.4) implicitly makes the simplifying assumption that the processes at different locations are independent of each other. We are further assuming that the intercepts are equal to zero and that

$$Y_1^l(t) = \theta_{11}^l(t) + v_1^l(t), \quad \theta_{11}^l(2) = \theta_{11}^l(1) + w_{11}^l(2). \tag{4.4.5}$$

The errors $v_i^l(t)$, $w_{ji}^l(2)$ are assumed by the collective to be mutually independent of each other following a Gaussian distribution with mean zero and known variance $V_i^l(t)$ and $W_{ji}^l(2)$ respectively. Assume further that each panel has provided prior information about the parameter vector at time $t = 1$, such that a_{ji}^l is the mean prior estimate of $\theta_{ji}^l(1)$, whilst its variance is elicited to be c_{ji}^l.

Now assume that the collective has agreed on a linear utility factorisation over the attributes and that each panel has individually agreed to model every individual conditional utility function as a cubic. Let $r_i^l(t) = y_i^l(t)$ and assume that decisions are not arguments of the utility function. We let

$$u^{\mathcal{G}}(\cdot) = \sum_{i \in [4]_1} \sum_{t \in [2]} \sum_{l \in [r]} -\gamma_i(t) y_i^l(t)^3, \tag{4.4.6}$$

where $\gamma_i(t) \in \mathbb{R}_{>0}$. It is easy to deduce by comparing equations (4.1.3) and (4.4.6) that this factorisation is a member of the class of compatible utilities. Furthermore the utility function, for each attribute of the decision problem, is assumed to be the same for the r geographical locations and that the overall score of an attribute is equal to the sum of the scores of each region for that attribute. The cubic utility function is a member of the family of constant relative risk aversion utilities, used in the literature to model risk aversion [see Wakker, 2008].

Now that the IDSS has been fully defined for this example, we can show how the algorithm works symbolically when the overarching structure is the LMDM. Recall that the third moment of a Gaussian distribution with mean μ and variance ψ is equal to $\mu^3 + 3\mu\psi$. Since this is a function of the first two moments (as for any high-order moment of a Gaussian, see Appendix C.2.1), the algorithm again consists of a sequential use of the tower property for the first two conditional moments. This starts with panels G_3

$$h_3(1) = -\sum_{l\in[r]}\left(\sum_{i=2}^{3}\gamma_i(1)y_i^l(1)^3 + 3\gamma_3(2)\mathbb{E}(\theta_{13}^l(2)^2\theta_{23}^l(2)\theta_{12}^l(2)Y_1^l(2)^3)\right)$$

$$h_3(2) = -\gamma_3(2)\sum_{l\in[r]}\left(3V_3^l(2)\mathbb{E}((\theta_{13}^l(2) + \theta_{23}^l(2)\theta_{12}^l(2))Y_1^l(2)) + \mathbb{E}((\theta_{13}^l(2)Y_1^l(2))^3)\right)$$

$$h_3(3) = -\sum_{l\in[r]}\left(\gamma_2(2) + \gamma_3(2)\mathbb{E}(\theta_{23}^l(2)^3)\right)\mathbb{E}(\theta_{12}^l(2)^3Y_1^l(2)^3)$$

$$h_3(4) = -\sum_{l\in[r]}\left(\gamma_2(2) + \gamma_3(2)\mathbb{E}(\theta_{23}^l(2)^3)\right)3V_2^l(2)\mathbb{E}(\theta_{12}^l(2)Y_1^l(2))$$

$$h_3(5) = -3\gamma_3(2)\sum_{l\in[r]}\mathbb{E}(\theta_{13}^l(2)\theta_{23}^l(2)^2Y_1^l(2))\left(\mathbb{E}(\theta_{12}^l(2)^2Y_1^l(2)^2) + \mathbb{V}(\theta_{12}^l(2)Y_1^l(2))\right)$$

$$h_3(6) = -3\gamma_3(2)\sum_{l\in[r]}\mathbb{E}(\theta_{13}^l(2)\theta_{23}^l(2)^2Y_1^l(2))V_2^l(2)$$

TABLE 4.7: Definition of the terms $h_2(i)$ in $\bar{u}_{2,2}$.

and G_4 computing, respectively, $\bar{u}_{2,3}$ and $\bar{u}_{2,4}$. Specifically, we have that

$$\bar{u}_{2,4} = -\sum_{l\in[r]}\left(\gamma_4(1)y_4^l(1)^3 + \gamma_4(2)\left(\mathbb{E}(\theta_{14}^l(2)^3Y_1^l(2)^3) + 3V_4^l(2)\mathbb{E}(\theta_{14}^l(2)Y_1^l(2))\right)\right), \quad (4.4.7)$$

$$\bar{u}_{2,3} = -\sum_{l\in[r]}\left(\gamma_3(1)y_3^l(1)^3 + \gamma_2(1)y_2^l(1)^3 + \gamma_2(2)y_2^l(2)^3 + \gamma_3(2)h_3\right), \quad (4.4.8)$$

where $h_3 = \mathbb{E}(\mathbb{E}(Y_3^l(2)^3 \mid \cdot)) + 3V_3^l(2)\mathbb{E}(\mathbb{E}(Y_3^l(2) \mid \cdot))$ and

$$\mathbb{E}(Y_3^l(2) \mid \cdot) = \theta_{13}^l(2)Y_1^l(2) + \theta_{23}^l(2)Y_2^l(2), \quad (4.4.9)$$

$$\mathbb{E}(Y_3^l(2)^3 \mid \cdot) = \sum_{i\in[2]}(\theta_{i3}^l(2)Y_i^l(2))^3 + 3\sum_{\substack{j+k=3\\j\neq k\in\mathbb{Z}_{\geq1}}}(\theta_{j3}^l(2)Y_j^l(2))^2\theta_{k3}^l(2)Y_k^l(2). \quad (4.4.10)$$

Now, equations (4.4.8)-(4.4.10) are functions of $Y_2^l(2)$ and $Y_2^l(1)$ only and can therefore $\longrightarrow G_2$. Then panel $G_2 : \bar{u}_{2,2}$ which can be written as the sum of the terms $h_2(i)$, $i \in [6]$, defined in Table 4.7.

All the terms in Table 4.7, as well as the rhs of equation (4.4.7), are a function of $Y_1^l(2)$ only and can therefore be sent to panel G_1, which sums these two incoming messages. Panel G_1 then applies sequentially the tower rules to compute $\bar{u}_{2,1}$. This corresponds to the sum of the rhs of (4.4.7) and the terms in Table 4.7 where $Y_1^l(2)$ and $Y_1^l(2)^3$ are substituted with $\theta_{11}^l(2)$ and $\theta_{11}^l(2)^3 + 3\theta_{11}^l(2)V_1(2)$ respectively.

The algorithm then considers the first time slice. Because of the very regular structure of the LMDM the expressions resulting from the first time slice are identical to the ones at the second time point. Since in an LMDM the parameters are all independent of each other, the expected utility function can be deduced by simply computing the expectation

of each of these. Specifically, letting for $i \in [2]$ and $j \in [4]_1$

$$k_{11}^l(2) = \mathbb{E}(\theta_{11}^l(2)^3) = ((a_{11}^l)^3 + 3a_{11}^l(c_{11}^l + V_1^l(2) + W_{11}^l(2))),$$

$$k_{11}^l(1) = \mathbb{E}(\theta_{11}^l(1)^3) = ((a_{11}^l)^3 + 3a_{11}^l(c_{11}^l + V_1^l(1))),$$

$$k_{ij}^l(1) = \mathbb{E}(\theta_{ij}^l(1)^3) = ((a_{ij}^l)^3 + 3a_{ij}^l c_{ij}^l),$$

$$k_{ij}^l(2) = \mathbb{E}(\theta_{ij}^l(2)^3) = ((a_{ij}^l)^3 + 3a_{ij}^l(c_{ij}^l + W_{ij}^l(2))),$$

$$d_{ij}^l(2) = \mathbb{E}(\theta_{ij}^l(2)^2) = (a_{ij}^l)^2 + c_{ij}^l + W_{ij}^l(2),$$

$$d_{ij}^l(1) = \mathbb{E}(\theta_{ij}^l(1)^2) = (a_{ij}^l)^2 + c_{ij}^l,$$

we can write the expected utility function as

$$\bar{u}(\boldsymbol{d}) = \sum_{l \in [r]} \left(\sum_{t \in [2]} \left(\sum_{i \in [4]_1} \gamma_i(t)\bar{u}_i^l(t) + \gamma_3(t)(\bar{m}_3^l(t) + \bar{n}_3^l(t) + k_{23}^l(t)\bar{u}_2(t)) \right) \right) \qquad (4.4.11)$$

where

$$\bar{u}_i^l(1) = k_{1i}^l(1)k_{11}^l(1) + 3V_i^l(1)a_{1i}^l a_{11}^l, \quad \bar{n}_3^l(t) = 3a_{13}^l d_{23}^l(t)(d_{12}^l(t)k_{11}^l(t) + V_2^l(t)a_{11}^l),$$

$$\bar{u}_i^l(2) = k_{1i}^l(2)k_{11}^l(2) + 3V_i^l(2)a_{1i}^l a_{11}^l, \quad \bar{m}_3^l(t) = 3V_3^l(t)a_{23}^l a_{12}^l(a_{11}^l + k_{11}^l(t)d_{13}^l(t)).$$

There are a few important points to notice here:

- because of the form of the utility factorisation in equation (4.4.6), the expected utility consists of the sum of the expected scores at each location l and of the sum, at each of these locations, of the scores associated to the two time slices. This result is a direct consequence of the independence of the processes at different locations. However, it would have been straightforward to embellish the example to allow the different processes to be dependent on one another, by defining a hierarchical model over $\boldsymbol{\theta}(t)^{\mathrm{T}} = (\boldsymbol{\theta}_i(t)^{\mathrm{T}})_{i \in [4]}$, where $\boldsymbol{\theta}_i(t)$ is the vector including the parameters $\theta_{ji}(t)$ in equations (4.4.4) and (4.4.5), $i \in [4]$;

- the expected utility in equation (4.4.11) is a polynomial, where the unknown quantities are the individual judgements delivered by the panels. This polynomial has for this example degree six and it is not a simple multilinear combination of the unknowns. Note that knowing the shape of the expected utility allows potential decision centres to understand how different factors influence the decision making process: we focus on these issues in Chapter 5;

- the IDSS would have not been able to compute the expected utility of equation (4.4.11) if the panels had only delivered the mean estimates of the variables under

their jurisdiction. The quantities $V_i^l(t)$, $W_{ij}^l(2)$ and c_{ij}^l represent levels of uncertainty concerning these mean estimates. If we just plug in the expectations of relevant means and set equal to zero the above measures of uncertainty we obtain a spurious evaluation of an expected utility score corresponding to

$$\bar{u}(\boldsymbol{d}) = \sum_{l\in[r]}\sum_{t\in[2]}\left(\sum_{i\in[4]_1}\gamma_i(t)(a_{11}^l)^3(a_{1i}^l)^3 + \gamma_3^t(a_{11}^l)^3(a_{23}^l)^2(a_{12}^l)^2(3a_{13}^l + a_{23}^l a_{12}^l)\right).$$

(4.4.12)

Equation (4.4.12) is way different from (4.4.11). A DSS that provides expected utility scores from equation (4.4.12) could thus lead decision centres to behave as non expected utility maximisers and put them in danger of adopting indefensible countermeasures.

4.4.5   A Network of Dynamic Models

The last example we present in this chapter consists of a sequence of generic models, whose input/output relationships respect the DAG in Figure 2.11 consisting of a Markov chain structure. We assume again the IDSS needs to identify the expected utility score associated to a specific policy, therefore skipping the optimisation steps, and suppress the dependence on the decisions.

As before, assume the variables $Y_i(t)$, $t \in [T]$, $i \in [4]$, are observed at each time point at r different locations in space and let $\boldsymbol{Y}_i(t) = (Y_i^1(t), \ldots, Y_i^r(t))^{\mathrm{T}}$, where $Y_i^l(t)$ denotes the i-th variable at time t and location l. Further assume that the compatible utility factorisation is linear over time, space and nodes of the graph, i.e.

$$u^{\mathcal{G}} = \sum_{i\in[4]}\sum_{t\in[T]}\sum_{l\in[r]} k_{til}u_{til}(y_i^l(t)),$$

Suppose further each panel individually agreed to model its marginal utility using a simple quadratic function of the form

$$u_{til} = -\gamma_i(t)y_i^l(t)^2.$$

Assume panel G_1 agreed to use a complex sampling computational algorithm delivering as output the first two moments of the distribution of $\boldsymbol{Y}_1(t)$, $t \in [T]$. Let $\boldsymbol{a}_1(t)$ and $C_1(t)$ the mean and covariance matrix respectively of $\boldsymbol{Y}_1(t)$ deduced from the agreed algorithm, with $c_1^{p,s}(t)$ the entry in position (p, s) of $C_1(t)$. Let also $a_1^l(t)$ and $c_1^l(t)$ be the mean and the variance of $Y_1^l(t)$ and $\boldsymbol{c}_1(t) = (c_1^1(t), \ldots, c_1^r(t))^{\mathrm{T}}$.

Panel G_2 agrees to model $\boldsymbol{Y}_2(t)$ according to r distinct MDMs defined by the following observation equations

$$Y_2^l(t) = \boldsymbol{Y}_1(t)^{\mathrm{T}} \boldsymbol{\theta}_2^l(t) + v_2^l(t), \qquad v_2^l(t) \sim (0, V_2^l(t)),$$

where $\boldsymbol{\theta}_2^l(t)$ is a r-dimensional vector of unknown parameters and $\mathbb{E}(V_2^l(t)) = b_2^l(t)$. The system equations are defined as

$$\boldsymbol{\theta}_2^l(t) = G_2^l(t) \boldsymbol{\theta}_2^l(t-1) + \boldsymbol{w}_2^l(t), \qquad \boldsymbol{w}_2^l(t) \sim (\boldsymbol{0}, W_2^l(t)),$$

where $G_2^l(t)$ is a $r \times r$ matrix having known entries and $W_2^l(t)$ is a diagonal matrix with known entries $w_{2,l}^{i,j}$ in position (i,j). Assume that is given a prior distribution with mean $\boldsymbol{a}_2^l$ and covariance C_2^l to $\boldsymbol{\theta}_2^l(1)$, where C_2^l is a diagonal matrix with $c_{2,l}^{i,j}$ in entry (i,j).

Panel G_3 agreed to use a simple LMDM, defined, for $l \in [r]$, by the following observation equations

$$Y_3^l(t) = \theta_{23}^l(t) Y_2^l(t) + v_3^l(t),$$

and system equations

$$\theta_{23}^l(t) = \theta_{23}^l(t-1) + w_{23}^l(t).$$

The errors $v_t^l(3)$ and $w_t^l(2,3)$ are assumed by G_3 to be mutually independent with mean zero and variance $V_3^l(t)$ and $W_{23}^l(t)$ respectively. These variances are assumed to be unknown, but the panel provides a prior mean $b_3^l(t)$ for $V_3^l(t)$ and $r_{23}^l(t)$ for $W_{23}^l(t)$. Prior means and variances are delivered also for the parameters $\theta_{23}^l(1)$ and denoted as a_{23}^l and c_{23}^l.

Lastly panel G_4 agreed to model the variables under their jurisdiction using a complex deterministic simulator for each region and to introduce uncertainty by defining a non-dynamic emulator over its outputs (see Section 2.3.5). The inputs of each simulator are $y_3^l(t)$ together with other known constants, whilst its output is $y_4^l(t)$. The emulator is then formally defined as

$$Y_4^l(t) = m(Y_3^l(t)) + e(Y_3^l(t)),$$

where $m(Y_3^l(t)) = \theta_{04}^l + \theta_{34}^l Y_3^l(t)$ and $e(Y_3^l(t))$ is a zero mean Gaussian process with covariance function $c^l(\cdot, \cdot) = W_4^l r^l(Y_3^l(t) - Y_3^l(t)')$, where r^l is a stationary correlation function such that $r^l(0) = 1$ For the purpose of this example, it is actually not necessary to specify such correlation function and we envisage the panel to have agreed on one such function [see e.g. Kennedy and O'Hagan, 2001]. The panel further agrees to deliver the following beliefs:

$$\mathbb{E}(\theta_{04}^l) = a_{04}^l, \quad \mathbb{E}(\theta_{34}^l) = a_{34}^l, \quad \mathbb{V}(\theta_{04}^l) = c_{04}^l, \quad \mathbb{V}(\theta_{34}^l) = c_{34}^l, \quad \mathbb{E}(W_4^l) = r_4^l.$$
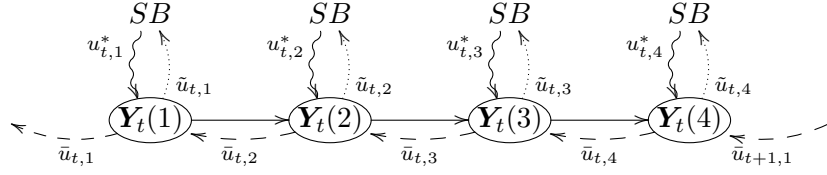
FIGURE 4.6: Description of the propagation algorithm over the directed acyclic graph of Figure 2.11.

We now show how the algorithm works symbolically under the overarching structure defined above. Its steps are summarised in Figure 4.6 where the arrows have the same meaning as in Example 4.3.1. For conciseness here we assume that the time horizon is equal to two and we start illustrating the algorithm from panel G_4 overseeing $\boldsymbol{Y}_2(4)$. Note that by construction $\tilde{u}_{2,4} = u^{\mathcal{G}}$. Letting $\dot{u}_{2,4} = u^{\mathcal{G}} - \sum_{l \in [r]} k_{24l} u_{24l}$, panel G_4 computes $\bar{u}_{2,4}$, as

$$
\begin{aligned}
\bar{u}_{2,4} &= \dot{u}_{2,4} - \sum_{l \in [r]} k_{24l} \gamma_4^l(2) \mathbb{E}(Y_4^l(2)^2) \\
&= \dot{u}_{2,4} - \sum_{l \in [r]} k_{24l} \gamma_4^l(2)(\mathbb{E}(Y_4^l(2))^2 + \mathbb{V}(Y_4^l(2))) \\
&= \dot{u}_{2,4} - \sum_{l \in [r]} \gamma_4^l(2) k_{24l} \left[\tau_4^l + 2a_{04}^l a_{34}^l \mathbb{E}(Y_3^l(2)) + \tau_{34}^l \mathbb{E}(Y_3^l(2)^2) \right],
\end{aligned}
$$

where $\tau_{04}^l = {a_{04}^l}^2 + r_4^l + c_{04}^l$ and $\tau_{34}^l = {a_{34}^l}^2 + c_{34}^l$.

This function is then $\longrightarrow G_3$. Note that because of the topology of the DAG, $\tilde{u}_{2,3} = \bar{u}_{2,4}$. Let $h_{2,3}^l = -2\gamma_4^l(2) k_{24l} a_{04}^l a_{34}^l$ and

$$
\begin{aligned}
\dot{u}_{2,3} &= \sum_{i \in [4]} \sum_{l \in [r]} k_{1il} u_{1il} + k_{21l} u_{21l} + k_{22l} u_{22l} - \gamma_4^l(2) k_{24l} \tau_4^l), \\
g_{2,3}^l &= -(\gamma_4^l(2) k_{24l} \tau_{34}^l + k_{23l} \gamma_3^l(2)).
\end{aligned}
$$

Then, $\bar{u}_{2,3}$ is equal to

$$
\bar{u}_{2,3} = \dot{u}_{2,3} - \sum_{l \in [r]} g_{2,3}^l \mathbb{E}(Y_3^l(2)^2) + h_{2,3}^l \mathbb{E}(Y_3^l(2)),
$$

where $\mathbb{E}(Y_3^l(2)) = a_{23}^l \mathbb{E}(Y_2^l(2))$ and $\mathbb{E}(Y_3^l(2)^2) = ({a_{23}^l}^2 + r_{23}^l(2) + c_{23}^l) \mathbb{E}(Y_2^l(2)^2) + b_3^l(2)$.

Rearranging and noting that $\tilde{u}_{2,2} = \bar{u}_{2,3}$, we have that

$$
\tilde{u}_{2,2} = \dot{u}_{2,2} - \sum_{l \in [r]} \left[[g_{2,3}^l({a_{23}^l}^2 + r_{23}^l(2) + c_{23}^l) + k_{22l} \gamma_2^l(2)] \mathbb{E}(Y_2^l(2)^2) + h_{2,3}^l a_{23}^l \mathbb{E}(Y_2^l(2)) \right],
$$

$$
(4.4.13)
$$

where $\bar{\gamma}_i^l(t) = k_{til}\gamma_i^l(t)$, $\tau_1^l(t) = a_1^l(t)^2 + c_1^l(t)$, $\bar{u}_3^l(1) = c_{23}^l + (a_{23}^l)^2$, $\bar{u}_3^l(2) = \bar{u}_3^l(1) + r_{23}^l(2)$ and

$$
\begin{aligned}
\tau_{mix}^l(2) &= b_3^l(2)\bar{\gamma}_3^l(2) + \bar{\gamma}_4^l(2)\tau_4^l + 2a_{04}^l a_{34}^l a_{23}^l \boldsymbol{a}_1(2)^{\mathrm{T}} G_2^l(2)\boldsymbol{a}_2^l, \\
\tau_{mix}^l(1) &= b_3^l(1)\bar{\gamma}_3^l(1) + \bar{\gamma}_4^l(1)\tau_4^l + 2a_{04}^l a_{34}^l a_{23}^l \boldsymbol{a}_1(1)^{\mathrm{T}} \boldsymbol{a}_2^l, \\
\bar{u}_2^l(2) &= {\boldsymbol{a}_2^l}^{\mathrm{T}} G_2^l(2)^{\mathrm{T}}(C_1(2) + \boldsymbol{a}_1(2)\boldsymbol{a}_1(2)^{\mathrm{T}}) G_2^l(2)\boldsymbol{a}_2^l + b_2^l(2) \\
&\quad + \sum_{i,j \in [r]} (A^{i,j} + w_{2,l}^{i,j})(a_1^i(2)a_1^j(2) + c_1^{i,j}(2)) \\
\bar{u}_2^l(1) &= {\boldsymbol{a}_2^l}^{\mathrm{T}}(C_1(2) + \boldsymbol{a}_1(2)\boldsymbol{a}_1(2)^{\mathrm{T}})\boldsymbol{a}_2^l + b_2^l(1) + \sum_{i,j \in [r]} c_{2,l}^{i,j}(a_1^i(1)a_1^j(1) + s_1^{i,j}(1)),
\end{aligned}
$$

with $A = G_2^l(2)C_2^l G_2^l(2)^{\mathrm{T}}$, such that $A^{i,j}$ is its entry in position (i,j).

4.5 Conclusions

Having formally defined in the previous chapter the IDSSs and discussed the conditions for sound group inferences, we introduced here new distributed propagation algorithms for the exact computation of an IDSS expected utility scores. These work for a novel asymmetric dynamic decision problem class, which embeds partial factorisations of utility functions. These types of factorisations have not been considered in expected utility computations before.

As shown by the above examples, all calculations are straightforward and scale up, albeit with a large number of moments or probabilities to be computed, stored and transmitted between panels. However, these quantities can be provided by an IDSS. So the large number of computations necessary for coherently evaluating different policies are actually trivial ones and operationally computable in real time. We note that the algorithms we define in these multi-ESs are closely related to the ones already cited for the propagation of probabilities and expected utilities in graphical structures, which have now been successfully implemented in many large applications. So we can be confident that our methods remain feasible for current and much larger applications.

Chapter 5

The Algebra of Integrating Partial Belief Systems

The examples of Section 4.4 showed that the IDSS's expected utilities are often polynomial functions whose indeterminates are individually delivered by different panels. By requesting only this information, the implementation of an IDSS can become orders of magnitude more manageable. This is because panels just need to communicate a few summaries from their sample: a trivial and fast task to perform.

Therefore, for the sole purpose of decision support, the full inferential conditions guaranteeing a sound and distributed analysis of Chapter 3 are actually too strict. These can be relaxed whilst retaining the coherence of the IDSS. Taking an algebraic approach, in this chapter we develop a methodology that ensures coherence in these types of partially defined systems, meaning systems where panels only deliver certain selected summaries. Following Leonelli et al. [2015b], we first define the new notion of algebraic expected utility of an IDSS. This definition enables us to impose new conditions tailored to the needs of an IDSS concerning the uncorrelatedness of certain polynomials. These conditions, often weaker than panel independence, can guarantee an IDSS is adequate. We discuss how and when it is possible to achieve adequacy in a number of typical frameworks. Importantly, this algebraic approach enables us to extend results about the computation of moments of decomposable functions in DAGs [Cowell et al., 1999, Nilsson, 2001] for a particular subclass of BN models.

Although this is not often recognised in the literature, algebraic approaches are intimately linked to symbolic ones, where probabilities are treated as indeterminates in a computer algebra system, thus not requiring an exact numerical specification. Symbolic inference is often used in sensitivity analysis to identify the probabilities a DM needs to be particularly precise about in her specification, as these might drastically change

the ranking of the available policies [French and West, 2003]. We symbolically define two important instances of models that can be part of an IDSS structural consensus, namely IDs and staged trees. We develop new symbolic inferential techniques for these two models, extending current literature to both decision and asymmetric problems. An extensive discussion of these results can be found in Leonelli et al. [2015a] and Görgen et al. [2015].

The structure of the chapter is as follows. In Section 5.1 we review the main developments in symbolic and algebraic inferential methods. In Section 5.2 we define algebraically the expected utility of an IDSS. Section 5.3 introduces polynomial conditions the panels' summaries need to entertain and we then show in Section 5.4 how these are able to guarantee adequacy and distributivity. Section 5.5 presents a variety of examples of these methods. In Section 5.6 we review recent symbolic methods for two particular integrating structures: IDs and staged trees. We conclude with a discussion.

5.1   A Brief Overview of Symbolic and Algebraic Methods in Statistics

As mentioned in the introduction, symbolic and algebraic methods for statistical inference have been developed independently by two different community of researchers. Symbolic ones have been the focus of mostly computer scientists, whilst algebraic ones of mathematicians and statisticians. However, in both cases, the probabilities associated to a statistical model are described as polynomials, whose structure is then exploited to study the inferential properties of such models in a variety of applications.

Symbolic methods focused mostly on graphical models and specifically on BNs, both discrete [Castillo et al., 1995, 1997a] and Gaussian [Castillo et al., 1997b]. Probabilities and moments associated to vertices are thought of as indeterminates of a computer algebra system (such as Maple or Matlab), where the model's probabilities can then be seen as polynomials with such indeterminates. Castillo et al. [1996] and Castillo and Kjærulff [2003] developed and implemented algorithms for the computation of polynomials associated to the joint probabilities of a BN model. In these works the bounds the probabilities of the vertices of the network have to respect were also identified, and these can then be used in sensitivity analyses. The fundamental idea behind this approach lies in the fact that once the qualitative structure of the model is constructed, numerical specifications of the involved probabilities do not need to be delivered a priori for the algorithms to run. Furthermore the above mentioned bounds inform a DM about probability values that cannot be delivered, thus simplifying the elicitation process.

However, the symbolic inferential algorithms of Castillo [1997] had several drawbacks. Most importantly these computed *every* possible monomial deriving from the symbolic definition of the model, only later dropping those associated to non compatible instantiations. This required the computation of the Cartesian product of the sets of probabilities associated to the vertices of the BN, which can become infeasible as the size of the network grows. In Darwiche [2003] a new symbolic inferential algorithm speeded up the computation of the probability polynomials, by converting the underlying BN into an *Arithmetic Circuit (AC)*. This circuit can then be evaluated through fast routines to output the required polynomials. Importantly, Darwiche [2003] also showed that a large number of probabilistic queries can be answered by computing certain *derivatives* of the probability polynomials, which are then easier to handle because the associated AC has a smaller dimension. Such methods were subsequently extended to DBNs [Brandherm and Jameson, 2004].

Beside the possibility of not having to elicit probabilities before any inference takes place, another great advantage of symbolic approaches is that these methods straightforwardly support various *sensitivity analyses*. In particular users can simply plug-in different numbers into the probability indeterminates and observe how these change the inferential results on a set of goal variables. Alternatively, much more sophisticated sensitivity techniques have now been implemented [Chan and Darwiche, 2001, 2004].

Although the computing capability of modern computers is constantly growing, it has been possible to apply symbolic approaches only in small/medium sized applications. This is because each probability specification is treated as an unknown variable in a computer system. This requires much more memory space than a single number. Furthermore, as shown by Cooper [1990], the number of monomials in BN models grows exponentially with the number of vertices, often making the computation of such polynomials prohibitive. However, we note here that the technology of IDSSs allows us to apply symbolic methods to much larger models. This is because we can define symbolically only the *outputs* of each component DSS, which is the only relevant information an IDSS needs to process. At the same time, this symbolic definition leads straightforwardly to the use of sensitivity analyses for the outputs of IDSSs, which can be extremely useful for building consensus among the panel members.

The alternative stream of research using polynomial techniques in statistics is usually referred to as *algebraic statistics*. Among others, this uses techniques from algebraic geometry and computational commutative algebra to gain insights into the structure of certain statistical models [Riccomagno, 2009]. The first developments of algebraic statistics were on the use of both Gröbner bases in Markov chain sampling algorithms

[Diaconis and Sturmfels, 1998] and commutative algebraic techniques in the design and analysis of experiments [Pistone and Wynn, 1996].

Subsequently, algebraic statistics also started to be applied to more general statistical models. Certain statistical models have been identified with certain algebraic varieties [Pistone et al., 2002]. Big part of the research in this area then focused on graphical models, as for example BNs [Garcia et al., 2005, Sullivant, 2008], MNs [Geiger et al., 2006] and trees [Settimi and Smith, 2000]. The number of applications of algebraic methods in statistics is now enormous [see e.g. the monographs, Drton et al., 2009, Gibilisco et al., 2010, Pachter and Sturmfels, 2005, Pistone et al., 2002].

To our knowledge, the methods we develop in the following sections are the first application of symbolic and algebraic techniques to Bayesian decision analysis.

5.2 An Algebraic Description of Integrating Systems

The computation of the expected utilities in the examples of Section 4.4 showed the difficulty of identifying the beliefs that need to be delivered to the IDSS by panels. The situation becomes even more complicated when the utility consensus includes multilinear or multiplicative utility factorisations introduced in Definitions 2.4.12 and 2.4.14 respectively (unless all panels can simply define independent marginal models for the variables under their jurisdiction).

Approaching the problem from an algebraic viewpoint allows us to identify the necessary panels' summaries and the required assumptions for adequacy. In order to do this we first need to define the expected utility polynomials. Recall that there are m panels of experts $\{G_i : i \in [m]\}$, $[m] = \{1, \ldots, m\}$, each delivering beliefs about $\boldsymbol{\theta}_i$, $i \in [m]$, parametrising the density of $\boldsymbol{Y}_i \mid \boldsymbol{d}$, for a decision $\boldsymbol{d} \in \mathcal{D}$. The conditional expected utility is $\bar{u}(\boldsymbol{d} \mid \boldsymbol{\theta})$, whilst $\bar{u}(\boldsymbol{d})$ is the expected utility.

Definition 5.2.1. *The conditional expected utility $\bar{u}(\boldsymbol{d} \mid \boldsymbol{\theta})$ of an IDSS is called **algebraic** if, for each $\boldsymbol{d} \in \mathcal{D}$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\bar{u}(\boldsymbol{d} \mid \boldsymbol{\theta})$ is a square-free polynomial*

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

in functions $\boldsymbol{\lambda}_i(\boldsymbol{\theta}_i, \boldsymbol{d})$ of the parameters $\boldsymbol{\theta}_i$ of panel G_i, $i \in [m]$.

We now explicitly define the algebraic conditional expected utilities $f_{\boldsymbol{d}}$. Let $\boldsymbol{\lambda}_i(\boldsymbol{\theta}_i, \boldsymbol{d}) = (\lambda_{ji}(\boldsymbol{\theta}_i, \boldsymbol{d}))_{j \in [s_i]}^{\mathrm{T}}$, for an $s_i \in \mathbb{Z}_{\geq 1}$, and $\boldsymbol{b} \in B = \bigtimes_{i \in [m]} B_i$, where $B_i = [s_i]^0 = [s_i] \cup \{0\}$, $i \in [m]$. For a given $\boldsymbol{b}$, let $i \in B_j$ be its j-th entry and let $b(j, i) = 0$ if $j \neq i$, $b(j, i) = 1$

if $j = i$ and $b(0, i) = 1$, for $i \in [m]$ and $j \in [s_i]$. Then, if we define $\lambda_{0i}(\boldsymbol{\theta}_i, \boldsymbol{d}) = 1$, for every $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i$, $\boldsymbol{d} \in \mathcal{D}$ and $i \in [m]$, we can write

$$f_{\boldsymbol{d}}\left(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, \boldsymbol{d}), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, \boldsymbol{d})\right) = \sum_{\boldsymbol{b} \in B} k_{\boldsymbol{b}, \boldsymbol{d}} \lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, \boldsymbol{d}), \tag{5.2.1}$$

where

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

is a monomial, having at most one term not unity delivered by each panel. So in particular, $f_{\boldsymbol{d}}$ is a square-free polynomial. For a given $\boldsymbol{b} \in B$, let

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

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

Definition 5.2.2. *Call an IDSS **score separable** if, in the notation above, the collective agrees that, for all decisions $\boldsymbol{d} \in \mathcal{D}$ and all indices $\boldsymbol{b} \in B$ such that $k_{\boldsymbol{b}, \boldsymbol{d}} \neq 0$,*

$$\mathbb{E}\left(\lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, \boldsymbol{d})\right) = \prod_{i \in [m]} \prod_{j \in [s_i]^0} \mu_{ji}(\boldsymbol{d}). \tag{5.2.2}$$

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

Lemma 5.2.3. *Suppose panel G_i delivers its vectors of expectations $\boldsymbol{\mu}_i(\boldsymbol{d})$, $i \in [m]$, $\boldsymbol{d} \in \mathcal{D}$, to the SB. Then, under the collective assumptions of an algebraic conditional expected utility, if the IDSS is score separable then it is adequate.*

Proof. The definition of adequacy in Definition 3.2.6 means in this framework that the expected utility $\bar{u}(\boldsymbol{d})$ is a function of $\mu_{ji}(\boldsymbol{d})$ and $k_{\boldsymbol{b}, \boldsymbol{d}}$ only, for $i \in [m]$, $j \in [s_i]^0$, $\boldsymbol{b} \in B$ and $\boldsymbol{d} \in \mathcal{D}$. Now note that

$$\bar{u}(\boldsymbol{d}) = \mathbb{E}\left(f_{\boldsymbol{d}}\left(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, \boldsymbol{d}), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, \boldsymbol{d})\right)\right) = \sum_{\boldsymbol{b} \in B} k_{\boldsymbol{b}, \boldsymbol{d}} \mathbb{E}\left(\lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, \boldsymbol{d})\right).$$

The definition of score separability in equation (5.2.2) then implies that

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

from which the lemma follows. $\square$

We can therefore deduce from Lemma 5.2.3 that adequacy is guaranteed whenever score separability holds, under the assumption of an algebraic conditional expected utility. 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.

5.3   Moment and Partial Independence

Equation (5.2.2) together with Lemma 5.2.3 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 *partial independence.*

Definition 5.3.1. *Let $f_{\boldsymbol{d}}(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, \boldsymbol{d}), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, \boldsymbol{d}))$ be the algebraic conditional expected utility of an IDSS. We say that an IDSS is **partially independent** if*

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

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 (see Appendix D).

Definition 5.3.2. *Let $\boldsymbol{\theta} = (\theta_i)_{i \in [n]}^{\mathrm{T}} \in \mathbb{R}^n$ be a parameter vector and $\boldsymbol{b} = (b_i)_{i \in [n]}^{\mathrm{T}} \in \mathbb{Z}_{\geq 0}^n$. We say that $\boldsymbol{\theta}$ entertains **moment independence** of order $\boldsymbol{b}$ if, $\forall\, \boldsymbol{b}' = (b_i')_{i \in [n]}^{\mathrm{T}} \leq_{lex} \boldsymbol{b}$,*

$$\mathbb{E}\left(\boldsymbol{\theta}^{\boldsymbol{b}'}\right) = \prod_{i \in [n]} \mathbb{E}\left(\theta_i^{b_i'}\right).$$

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 parametrisation, where the cumulant generating function for a product of independent random variables (defined as a random sum of independent realisations) 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. Consider for instance two parameters θ_1 and θ_2. Suppose a conditional expected utility is equal to $\theta_1^2 \theta_2^2$ and

that a moment independence of order $(2, 2)$ holds. Then,

$$\mathbb{E}\left(\theta_1^2 \theta_2^2\right) = \mathbb{E}\left(\theta_1^2\right)\mathbb{E}\left(\theta_2^2\right) = \mathbb{E}(\theta_1)^2\mathbb{E}(\theta_2)^2 + \mathbb{E}(\theta_1)^2\mathbb{V}(\theta_2) + \mathbb{E}(\theta_2)^2\mathbb{V}(\theta_1) + \mathbb{V}(\theta_1)\mathbb{V}(\theta_2).$$
(5.3.1)

The same expression is obtained when using sequentially the tower rule of expectations and the law of total variance of Proposition 2.1.6 under the assumption of independence of the two parameters above. 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 (5.3.1) to hold *does not require* θ_1 and θ_2 to be independent.

5.4 Adequacy in Partially Defined Systems

Given the definitions of new independence concepts tailored for IDSSs, we can now study in which cases adequacy holds. The following result easily follows from the definition of partial independence.

Lemma 5.4.1. *Suppose* $f_{\boldsymbol{d}}(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, \boldsymbol{d}), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, \boldsymbol{d}))$ *is the algebraic conditional expected utility of an IDSS and that partial independence is in the structural consensus. The IDSS is then adequate if panel* G_i *delivers the expectations* $\boldsymbol{u}_i(\boldsymbol{d})$, $i \in [m]$, $\boldsymbol{d} \in \mathcal{D}$.

Proof. This result easily follows by noting that partial independence implies score separability. Specifically,

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

The result then follows from Lemma 5.2.3. □

Assuming the conditional expected utility is a polynomial in the panels' parameters, then under a specific moment independence assumption we have the following.

Lemma 5.4.2. *Assume* $f_{\boldsymbol{d}}(\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, \boldsymbol{d}), \ldots, \boldsymbol{\lambda}_m(\boldsymbol{\theta}_m, \boldsymbol{d}))$ *is the algebraic conditional expected utility of an IDSS,* $\boldsymbol{\theta}_i = (\theta_{ji})_{j \in [s_i]}^{\mathrm{T}}$ *and* $\lambda_{ji}(\boldsymbol{\theta}_i, \boldsymbol{d}) = \boldsymbol{\theta}_i^{\boldsymbol{a}_{ji}}$, *with* $\boldsymbol{a}_{ji} \in \mathbb{Z}_{\geq 0}^{s_i}$, $i \in [m]$, $j \in [s_i]$. *Let* $\boldsymbol{a}_i^* = (a_{ji}^*)_{j \in [s_i]}^{\mathrm{T}}$, *where* a_{ji}^* *is the greatest element in* $\{a_{ji} : j \in [s_i]\}$, $i \in [m]$, *and let* $\boldsymbol{a}^{*\mathrm{T}} = (\boldsymbol{a}_i^{*\mathrm{T}})_{i \in [m]}$. *Let* $\boldsymbol{\theta} = (\boldsymbol{\theta}_i^{\mathrm{T}})_{i \in [m]}^{\mathrm{T}}$ *and assume the structural consensus includes the assumption that* $\boldsymbol{\theta}$ *entertains moment independence of order* $\boldsymbol{a}^*$. *Then the IDSS is adequate if panel* G_i *delivers the vectors of expectations* $\boldsymbol{u}_i(\boldsymbol{d})$, $i \in [m]$, $\boldsymbol{d} \in \mathcal{D}$.

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

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

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

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

and the result follows. $\qquad \square$

Both Lemma 5.4.1 and 5.4.2 start with the assumption of an algebraic conditional expected utility. This is the case for many of the examples we consider in Section 5.5. However, we are able to identify families of utility factorisations and statistical models that ensure the associated conditional expected utility is algebraic. We first introduce two families of utilities that factorise according to the construction of the panels.

Definition 5.4.3. *Let $\boldsymbol{Y}_i$ be the vector overseen by panel G_i, $i \in [m]$, and $\boldsymbol{R}_i$ be a function of $\boldsymbol{Y}_i$ and $\boldsymbol{d} \in \mathcal{D}$ only. A multilinear factorisation over $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m$ is called* **panel separable**.

Definition 5.4.4. *Under the conditions of Definition 5.4.3, an additive factorisation over $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_m$ is called* additive panel separable.

Both families of panel separable and additive panel separable utilities can be seen as a particular instance of a compatible utility factorisation in Definition 4.1.8.

The probabilistic model class we consider here is a generalisation of the DAG linear regressions in equation (2.3.4) to generic polynomial functions. Henceforth we call these a *polynomial Structural Equation Model (SEM)* [see e.g. Bollen and Lang, 1993, Ullman and Bentler, 2003, Wall and Amemiya, 2000]. SEMs are widely used, especially recently, in the causal literature [Pearl, 2000].

Definition 5.4.5. *Let $\boldsymbol{Y} = (Y_i)_{i \in [n]}^{\mathrm{T}}$ be a random vector. A* **polynomial SEM** *is defined by, for $i \in [n]$,*

$$
Y_i = \sum_{\boldsymbol{b}_i \in B_i} \theta_{i \boldsymbol{b}_i} \boldsymbol{Y}_{[i-1]}^{\boldsymbol{b}_i} + \varepsilon_i,
$$

where $B_i \subset \mathbb{Z}_{\geq 0}^{i-1}$, ε_i is an error with mean zero and variance ψ, and $\theta_{i\boldsymbol{b}_i}$ is a parameter, $i \in [n]$, $\boldsymbol{b}_i \in B_i$.

An alternative formulation of the model in Definition 5.4.5 in terms of distributions is,

$$Y_i \mid \boldsymbol{\theta}_i, \boldsymbol{Y}_{[i-1]} \sim \Big(\sum_{\boldsymbol{b}_i \in B_i} \theta_{i\boldsymbol{b}_i} \boldsymbol{Y}_{[i-1]}^{\boldsymbol{b}_i}, \psi_i \Big),$$

where $\boldsymbol{\theta}_i = (\psi_i, \theta_{i\boldsymbol{b}_i})_{\boldsymbol{b}_i \in B_i}^{\mathrm{T}}$ and $i \in [n]$.

Note that these models are suitable candidates for a CK-class since their definition is qualitative in nature.

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

Theorem 5.4.6. *Suppose $\boldsymbol{R} = \boldsymbol{Y}$ and assume panel G_i is responsible for Y_i, $i \in [n]$. Assume that the utility consensus of an IDSS includes a panel separable utility and that the structural consensus includes a polynomial SEM. Suppose each panel agreed to model its marginal utility as a polynomial utility function. Then if panels are partially independent, the IDSS is score separable.*

The proof of this result can be found in Appendix A.4.1.

Theorem 5.4.6 together with Lemma 5.2.3 can be thought of as an instance of the important Theorem 3.4.3, specifically addressing adequacy in partial belief systems. An IDSS whose CK-class embeds the assumptions of Theorem 5.4.6 is able to uniquely compute expected utility scores from the individual judgements of the panels.

Note that, by construction, the partial independence condition of Theorem 5.4.6 actually corresponds to a moment independence. The order of such an independence depends on the polynomial form of both the SEM and the utility function. In Section 5.5.2 we identify the order of the moment independence condition required for adequacy for a subclass of polynomial SEMs.

5.5 Examples

5.5.1 Discrete Models

5.5.1.1 Independence Binary Models. We begin with a rather obvious setting where a small number of summaries are sufficient to determine an expected utility maximising decision. Let the CK-class specify that $\boldsymbol{R} = \boldsymbol{Y} = (Y_i)_{i \in [n]}^{\mathrm{T}}$, where each variable Y_i is binary and overseen by panel G_i. Assume the CK-class includes the belief

that $\perp\!\!\!\perp_{i\in[n]} Y_i \mid \boldsymbol{\theta}, \boldsymbol{d}$, where $\boldsymbol{\theta} = (\theta_i)_{i\in[n]}^{\mathrm{T}}$ and that θ_i can be read as the probability $\mathbb{P}(Y_i = 1 \mid \theta_i, \boldsymbol{d})$, whatever decision $\boldsymbol{d} \in \mathcal{D}$ is made, $i \in [n]$. We have already discussed in Section 3.6.1.1 how independence models can be part of the structural consensus of a CK-class.

As in the example in Section 3.3.2, suppose each panel G_i delivers the set of beta distributions $\mathrm{Be}(p_i, q_i)$ for $\theta_i \mid \boldsymbol{d}$, $i \in [n]$. Assume further the CK-class includes utility factorisations of the form

$$u(y_1, \ldots, y_n) = \sum_{i\in[n]} k_i y_i + \sum_{i\in[n]} \sum_{j\in[n]_i} k_{ij} y_i y_j.$$

With no further assumptions, the conditional expected utility can be written as

$$\bar{u}(\boldsymbol{d} \mid \boldsymbol{\theta}) = \sum_{i\in[n]} k_i \lambda_i(\theta_i, \boldsymbol{d}) + \sum_{i\in[n]} \sum_{j\in[n]_i} k_{ij} \lambda_i(\theta_i, \boldsymbol{d}) \lambda_j(\theta_j, \boldsymbol{d}), \tag{5.5.1}$$

where $\lambda_i(\theta_i, \boldsymbol{d}) = \theta_i$, $[n]_i = \{i+1, \ldots, n\}$. Thus, equation (5.5.1) is an algebraic conditional expected utility. In this example partial independence then corresponds to moment independence of order $\mathbf{1}$, where $\mathbf{1}$ is a vector of dimension n with 1 in all its entries. In this case, partial independence is implied by panel independence. Furthermore all the monomials $\lambda_{\boldsymbol{b}}(\boldsymbol{\theta}, d)$ formally defined in equation (5.2.1) here are monomials of degree either one or two corresponding respectively to $\lambda_i(\theta_i, \boldsymbol{d})$ or $\lambda_i(\theta_i, \boldsymbol{d}) \lambda_j(\theta_j, \boldsymbol{d})$, for $i \in [n]$ and $j \in [n]_i$.

Now let $\mu_i = p_i(p_i + q_i)^{-1} = \mathbb{E}(\theta_i \mid \boldsymbol{d})$ and assume partial independence is in the CK-class. We can see that taking the expectation of equation (5.5.1) we have that

$$\bar{u}(\boldsymbol{d}) = \sum_{i\in[n]} k_i \mu_i + \sum_{i\in[n]} \sum_{j\in[n]_i} k_{ij} \mu_i \mu_j,$$

where $\bar{u}(d)$ is the expected utility. Thus this IDSS is adequate. Note that this IDSS does not need panel independence to uniquely compute expected utility scores, but only a simple moment independence.

5.5.1.2 Staged Trees. Consider now the staged tree in Figure 2.7 of Section 2.3.2.6, that we report again on page 150 for convenience. As discussed in Section 3.6.1.7, staged trees can be part of a coherent IDSS whenever panels oversee disjoint subsets of either the position set or the stage set. Recall that this tree has four stages (coinciding with its positions) $w_0 = \{v_0\}$, $w_1 = \{v_1\}$, $w_2 = \{v_2\}$ and $w_3 = \{v_3, v_4\}$. Suppose there are three panels G_1, G_2 and G_3 having responsibility over w_0, $\{w_1, w_2\}$ and w_3 respectively.
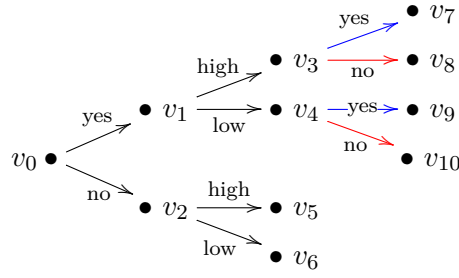
In the notation of Example 2.3.30, assume the collective has agreed on an additive utility factorisation such that

$$u(y_1, y_2, y_3) = k_1 u_1(y_1) + k_2 u_2(y_2) + k_3 u_3(y_3),$$

where Y_1, Y_2, Y_3 are the variables in the associated BN representation of this tree, and that they have been jointly able to further specify the criterion weights. Label the outcomes of this tree with a 1 for yes and high, whilst a 0 denotes no and low outcomes. Let $\psi_{ij} = u_i(j)$, where $j \in \{0, 1\}$, and recall that $\theta_{ss'} = \mathbb{P}(Y_s = s' \mid \boldsymbol{\theta}, \boldsymbol{d})$, for a situation $s \in S(\mathcal{T})$, $s' \in \mathcal{Y}_s = \{1, 0\}$ and $\boldsymbol{d} \in \mathcal{D}$, where $\boldsymbol{\theta}$ is the overall parameter vector. In this parametrisation, the staged tree in Figure 2.7 introduces the constraints $\theta_{31} = \theta_{41}$ and $\theta_{30} = \theta_{40}$, where for example $\theta_{31} = \mathbb{P}(Y_3 = \text{yes} \mid Y_2 = \text{high}, Y_1 = \text{yes}, \boldsymbol{\theta}, \boldsymbol{d})$.

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

$$\bar{u}(\boldsymbol{\theta} \mid \boldsymbol{d}) = \bar{u}_1 + \bar{u}_2 + \bar{u}_3, \tag{5.5.2}$$

where

$$
\begin{aligned}
\bar{u}_1 &= k_1 \psi_{11} \theta_{01} + k_1 \psi_{10} \theta_{00}, \\
\bar{u}_2 &= k_2 \psi_{21} \theta_{11} \theta_{01} + k_2 \psi_{20} \theta_{10} \theta_{01} + k_2 \psi_{21} \theta_{21} \theta_{00} + k_2 \psi_{20} \theta_{20} \theta_{00}, \\
\bar{u}_3 &= k_3 \psi_{31} \theta_{31} \theta_{11} \theta_{01} + k_3 \psi_{30} \theta_{30} \theta_{11} \theta_{01} + k_3 \psi_{31} \theta_{41} \theta_{10} \theta_{01} + k_3 \psi_{30} \theta_{40} \theta_{10} \theta_{01}.
\end{aligned}
\tag{5.5.3}
$$

So the conditional expected utility in equations (5.5.2)-(5.5.3) is again algebraic. The coefficients $k_{\boldsymbol{b},\boldsymbol{d}}$ of the monomials in equation (5.2.1) correspond to the jointly agreed criterion weights, and the unknown functions $\boldsymbol{\lambda}_i(\boldsymbol{\theta}_i, \boldsymbol{d})$ are as follows

$$
\begin{aligned}
\boldsymbol{\lambda}_1(\boldsymbol{\theta}_1, \boldsymbol{d}) &= (\psi_{11} \theta_{01}, \psi_{10} \theta_{00})^{\mathrm{T}}, \\
\boldsymbol{\lambda}_2(\boldsymbol{\theta}_2, \boldsymbol{d}) &= (\theta_{11}, \theta_{10}, \theta_{21}, \theta_{20}, \psi_{21} \theta_{11}, \psi_{20} \theta_{10}, \psi_{21} \theta_{21}, \psi_{20} \theta_{20})^{\mathrm{T}}, \\
\boldsymbol{\lambda}_3(\boldsymbol{\theta}_3, \boldsymbol{d}) &= (\psi_{31} \theta_{31}, \psi_{30} \theta_{30})^{\mathrm{T}}.
\end{aligned}
$$

Thus once more these polynomials are a simple multilinear function of probabilities delivered by different panels. Under partial independence, as guaranteed by Lemma 5.2.3, an IDSS so defined is adequate.

5.5.2 Bayesian Networks

We now consider the BN model class, where each variable of the network is defined as the DAG linear regression introduced in Definition 2.3.14. Note that this is an instance of the polynomial SEM of Definition 5.4.5. We refer to this model as a *linear SEM* over a DAG $\mathcal{G}$. Although such a model is often multivariate Gaussian, in general this does not have to be the case.

Just as Sullivant [2008], we consider regression parameters as indeterminates in a polynomial function. We associate these to edges and vertices of the underlying DAG. Let $\mathcal{G}$ be a DAG with vertex set $V(\mathcal{G}) = \{Y_i : i \in [n]\}$ and let, for $i \in [n]$, $\theta'_{0i} = \theta_{0i} + \varepsilon_i$ be the indeterminate associated to the vertex Y_i, whilst θ_{ij}, for $(Y_i, Y_j) \in E(\mathcal{G})$, is the indeterminate associated to the corresponding edge.[1] Define $\vec{P}_i$ as the set of rooted directed paths (see Appendix B) in $\mathcal{G}$ ending in Y_i. For every element $P \in \vec{P}_i$ we define $\boldsymbol{\theta}_P$ as

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

and we call $\boldsymbol{\theta}_P$ the *path monomial*.

Example 5.5.1. Consider the DAG in Figure 2.1. The set $\vec{P}_3$ for instance is equal to

$$\{(Y_3), (Y_2, (Y_2, Y_3)), (Y_1, (Y_1, Y_3)), (Y_1, (Y_1, Y_2), (Y_2, Y_3))\}, \tag{5.5.4}$$

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

For the purpose of this section 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. To illustrate this process, we present now an example of what we mean by an algebraic substitution.

Example 5.5.2. Consider again the DAG in Figure 2.1. For this network the variables are defined as

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

---

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

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

$$\begin{aligned}
Y_3 &= \theta_{03} + \theta_{13}(\theta_{01} + \varepsilon_1) + \theta_{23}(\theta_{02} + \theta_{12}Y_1 + \varepsilon_2) + \varepsilon_3 \\
&= \theta_{03'} + \theta_{13}\theta'_{01} + \theta_{23}\theta'_{02} + \theta_{23}\theta_{12}Y_1.
\end{aligned} \tag{5.5.5}$$

Note that an additional algebraic substitution can be performed in equation (5.5.5) so that

$$Y_3 = \theta'_{03} + \theta_{13}\theta'_{01} + \theta_{23}\theta'_{02} + \theta_{23}\theta_{12}\theta'_{01}. \tag{5.5.6}$$

It is of special interest that after this substitution Y_3 is now uniquely defined in equation (5.5.6) in terms of regression parameters. The following proposition formalises that this is in general the case for any variable of a DAG defined as a linear SEM.

Proposition 5.5.3. *Let $\mathcal{G}$ be a DAG with vertex set $\{Y_i : i \in [n]\}$, whose variables are defined as a linear SEM. Then through algebraic substitutions each variable Y_i, $i \in [n]$, 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. The variable Y_1 is by construction a root of the DAG $\mathcal{G}$. 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 construction and by the inductive hypothesis we have that

$$Y_n = \theta'_{0n} + \sum_{i \in \Pi_n} \theta_{in}Y_i = \theta'_{0n} + \sum_{i \in \Pi_n} \theta_{in} \sum_{P \in \vec{P}_i} \boldsymbol{\theta}_P. \tag{5.5.7}$$

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 (5.5.7). $\square$

Example 5.5.4. Equation (5.5.6) shows that Y_3 can be written as the sum of the monomials associated to each rooted path ending in Y_3, where the set of rooted paths is in equation (5.5.4).

An algebraic substitution corresponds to computing the conditional expectation of a random variable. This result is formalised by the following proposition.

Proposition 5.5.5. *Let $\boldsymbol{\theta}^{\mathrm{T}} = (\boldsymbol{\theta}_i^{\mathrm{T}})$, where $\boldsymbol{\theta}_i = (\theta'_{0i}, \theta_{ji})_{j \in \Pi_i}^{\mathrm{T}}$, $i \in [n]$. Under the conditions of Proposition 5.5.3, we have that*

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

Proof. This result is proven via the same inductive process as in the proof of Proposition 5.5.3. Note that the result holds for Y_1 since $\mathbb{E}(Y_1 \mid \boldsymbol{\theta}) = \theta'_{01}$. Assume the proposition is true for Y_{n-1} and consider Y_n. We have that

$$
\begin{aligned}
\mathbb{E}(Y_n \mid \boldsymbol{\theta}, \boldsymbol{d}) &= \mathbb{E}(\mathbb{E}(Y_n \mid \boldsymbol{\theta}_n, \boldsymbol{Y}_{\Pi_n}, \boldsymbol{d}) \mid \boldsymbol{\theta}, \boldsymbol{d}) = \mathbb{E}\left(\theta'_{0n} + \sum_{i \in [n]} \theta_{in} Y_i \mid \boldsymbol{\theta}, \boldsymbol{d}\right) \\
&= \theta'_{0n} + \sum_{i \in [n]} \theta_{in} \mathbb{E}(Y_i \mid \boldsymbol{\theta}, \boldsymbol{d}) = \theta'_{0n} + \sum_{i \in [n]} \theta_{in} \sum_{P \in \vec{P}_i} \boldsymbol{\theta}_P.
\end{aligned}
$$

Using a similar argument as in the proof of Proposition 5.5.3 the result follows. $\qquad\square$

5.5.2.1 Additive Factorisations. Given Propositions 5.5.3 and 5.5.5, we are now able to polynomially compute the expected utility of additive utility factorisations over the graphs in the case the marginal utility functions are polynomial.

Lemma 5.5.6. *For a linear SEM over a DAG $\mathcal{G}$ with vertex set $\{Y_i : i \in [n]\}$, suppose $\boldsymbol{R} = \boldsymbol{Y}$ and that the utility function $u(\boldsymbol{r})$ can be written*

$$
u(\boldsymbol{r}) = \sum_{i \in [n]} k_i u_i(y_i).
$$

Suppose further u_i is a polynomial utility function of degree n_i, i.e. $u_i = \sum_{j \in [n_i]} \rho_{ij} y_i^j$. Then the conditional expected utility is algebraic and can be written as

$$
\bar{u}(\boldsymbol{d} \mid \boldsymbol{\theta}) = \sum_{i \in [n]} k_i \sum_{j \in [n_i]^0} \rho_{ij} \sum_{|\boldsymbol{\alpha}_i| = j} \binom{j}{\boldsymbol{\alpha}_i} \boldsymbol{\theta}_{\mathcal{G}_i}^{\boldsymbol{\alpha}_i}, \tag{5.5.8}
$$

where $\boldsymbol{\alpha}_i = (\alpha_{ij})_{j \in [\#\vec{P}_i]}^{\mathrm{T}} \in \mathbb{Z}_{\geq 0}^{\#\vec{P}_i}$, $\#\vec{P}_i$ is the number of elements in $\vec{P}_i$, $|\boldsymbol{\alpha}_i| = \sum_{j \in [\#\vec{P}_i]} \alpha_{ij}$, $\boldsymbol{\theta}_{\mathcal{G}_i} = \prod_{P \in \vec{P}_i} \boldsymbol{\theta}_P$ and $\binom{j}{\boldsymbol{\alpha}_i}$ is a Multinomial coefficient (see Appendix C.1.2).

Proof. From Proposition 5.5.5, it follows that

$$
\mathbb{E}(u(\boldsymbol{R}) \mid \boldsymbol{\theta}, \boldsymbol{d}) = \sum_{i \in [n]} k_i \sum_{j \in [n_i]^0} \rho_{ij} \left(\sum_{P \in \vec{P}_i} \boldsymbol{\theta}_P\right)^j.
$$

The result follows applying the Multinomial theorem (see Appendix D). $\qquad\square$

Equation (5.5.8) is an instance of the computation of the moments of a decomposable function as studied in Cowell et al. [1999] and Nilsson [2001], where in this case we explicitly deduce the required monomials and their degree. In the following section we generalise their results to generic multilinear functions.

The proof of Lemma 5.5.6 provides an intuitive graphical description of the required monomials for the computation of the expected utility in equation (5.5.8). Furthermore the proof instructs us on how to compute any marginal moment of a linear SEM. Note that the j-th moment of any variable Y_i in such a BN can be written, via the Multinomial theorem, as the sum of the monomials $\boldsymbol{\theta}_{\mathcal{G}_i}$ with degree j. We can observe that, following from the properties of Multinomial coefficients, this sum can be thought of as the sum over the set of unordered j-tuples of paths ending in Y_i. Call this set $\vec{P}_i^j$. For a j-tuple $P \in \vec{P}_i^j$, the Multinomial coefficient in equation (5.5.8) counts the distinct permutations of the elements of P, denoted as n_P. We therefore have that

$$\sum_{|\boldsymbol{\alpha}_i|=j} \binom{j}{\boldsymbol{\alpha}_i} \boldsymbol{\theta}_{\mathcal{G}_i}^{\boldsymbol{\alpha}_i} = \sum_{P \in \vec{P}_i^j} n_P \prod_{p \in P} \boldsymbol{\theta}_p. \qquad (5.5.9)$$

Equation (5.5.9) is an intuitive graphical interpretation of equation (5.5.8).

Example 5.5.7. For simplicity we consider now the vertex Y_4 in the DAG of Figure 2.1. The set $\vec{P}_4$ is equal to $\{(Y_4), (Y_1, (Y_1, Y_4))\}$. From equation (5.5.8), Y_4^2 can be written as

$$\theta_{04}'^2 + \theta_{01}'^2 \theta_{14}^2 + 2\theta_{01}' \theta_{14} \theta_{04}'. \qquad (5.5.10)$$

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

$$\vec{P}_4^2 = \Big\{ \big((Y_4), (Y_4)\big), \big((Y_1, (Y_1, Y_4)), (Y_1, (Y_1, Y_4))\big), \big((Y_4), (Y_1, (Y_1, Y_4))\big) \Big\}.$$

The first and second monomial in equation (5.5.10) 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 (5.5.10)

From Lemma 5.5.6 we can deduce the independence needed for adequacy in BNs.

Theorem 5.5.8. *Suppose $\boldsymbol{R} = \boldsymbol{Y}$ and that the structural consensus of an IDSS includes a linear SEM over a DAG with vertex set $\{Y_i : i \in [n]\}$, where panel G_i oversees the vertex Y_i of $\mathcal{G}$, $i \in [n]$. Suppose the utility consensus consists of an additive panel separable utility function and panel G_i agreed to model its marginal utility with a polynomial utility function of degree $n_i \in \mathbb{Z}_{\geq 1}$, $i \in [n]$. Let l_P be the length of a rooted path P, $l_i = \sum_{P \in \vec{P}_i} l_P$ and $\boldsymbol{\alpha}_i \in \mathbb{Z}_{\geq 0}^{l_i}$. If $\boldsymbol{\theta}_{\mathcal{G}_i}$ entertains moment independence of order $\boldsymbol{\alpha}_i$ for every $|\boldsymbol{\alpha}_i| = n_i$ and $i \in [n]$, then the IDSS is score separable.*

Proof. Under the assumptions of the theorem, the conditional expected utility function can be written as in equation (5.5.8). From the linearity of the expectation operator we

have that

$$\bar{u}(\boldsymbol{d}) = \mathbb{E}(\bar{u}(\boldsymbol{d} \mid \boldsymbol{\theta})) = \sum_{i \in [n]} k_i \sum_{j \in [n_i]^0} \rho_{ij} \sum_{|\boldsymbol{\alpha}_i| = j} \binom{j}{\boldsymbol{\alpha}_i} \mathbb{E}\left(\boldsymbol{\theta}_{\mathcal{G}_i}^{\boldsymbol{\alpha}_i} \mid \boldsymbol{d}\right).$$

Noting that $\boldsymbol{\theta}_{\mathcal{G}_i} = \prod_{P \in \vec{P}_i} \boldsymbol{\theta}_P$, $\boldsymbol{\theta}_P = \prod_{Y_i \in P} \theta'_{0i} \prod_{(Y_j, Y_k) \in P} \theta_{jk}$ and applying moment independence, we have that

$$\bar{u}(\boldsymbol{d}) = \sum_{\substack{i \in [n], j \in [n_i]^0, \\ |\boldsymbol{\alpha}_i| = j}} k_i \rho_{ij} \binom{j}{\boldsymbol{\alpha}_i} \prod_{P \in \vec{P}_i, Y_l \in P} \mathbb{E}\left(\theta'^{\alpha_{il}}_{0l} \theta^{\alpha_{lCh_l}}_{lCh_l} \mid \boldsymbol{d}\right) \prod_{(Y_j, Y_k) \in P \setminus (Y_l, Y_{Ch_l})} \mathbb{E}\left(\theta^{\alpha_{ik}}_{jk} \mid \boldsymbol{d}\right),$$

where α_{ij} is the element of $\boldsymbol{\alpha}_i$ associated to θ_{kj}. The lemma then follows since $\bar{u}(\boldsymbol{d})$ is a multilinear function of expectations individually delivered by the panels. $\square$

Theorem 5.5.8 can be seen as an instance of Theorem 5.4.6 where we have been able to identify the specific moment independences necessary for the IDSS's adequacy. By requesting the collective to agree on these independences, the IDSS can then quickly produce a unique expected utility score for each policy. Furthermore, panels are informed about the summaries that they need to deliver to the IDSS since these are the only quantities the expected utility is a function of.

We can observe that in this case the functions λ_{ij} panel G_i delivers, $i \in [n]$, $j \in [s_i]$, are simply powers of monomials having one or two indeterminates only. Therefore, if panels deliver the expectation of these functions, where the required degrees are specified by Theorem 5.5.8, an IDSS so defined will be adequate.

5.5.2.2 Multilinear Factorisations. The algebraic approach we have taken in this chapter allows us to generalise straightforwardly the results on additive/decomposable functions to multilinear functions. We start analysing this more general case by stating two introductory results.

Proposition 5.5.9. *Let $\mathcal{G}$ be a DAG with vertex set $\{Y_i : i \in [n]\}$ and assume $\boldsymbol{R} = \boldsymbol{Y}$. Assume a multilinear utility factorisation over the graph such that*

$$u(\boldsymbol{r}) = \sum_{I \in \mathcal{P}_0([n])} k_I \prod_{i \in I} u_i(y_i),$$

where $\mathcal{P}_0$ denotes the power set without the empty set. Then letting each marginal utility $u_i(y_i)$ be polynomial of degree n_i and $\boldsymbol{n} = (n_i)_{i \in [n]}^{\mathrm{T}} \in \mathbb{Z}^n$,

$$u(\boldsymbol{r}) = \sum_{\boldsymbol{0} <_{lex} \boldsymbol{\alpha} \leq_{lex} \boldsymbol{n}} c_{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\alpha}}, \tag{5.5.11}$$

where $\boldsymbol{\alpha} = (\alpha_i)_{i\in[n]}^{\mathrm{T}} \in \mathbb{Z}_{\geq 0}^n$, $c_{\boldsymbol{\alpha}} = k_J \prod_{j\in J} \rho_{j\alpha_j}$ *and* $J = \{j \in [n] : \alpha_j \neq 0\}$.

The proof of this result is given in Appendix A.4.2.

Proposition 5.5.10. *For a linear SEM over $\mathcal{G}$ with vertex set $\{Y_i : i \in [n]\}$, letting* $\boldsymbol{\alpha} = (\alpha_i)_{i\in[n]}^{\mathrm{T}} \in \mathbb{Z}_{\geq 0}^n$ *and assuming $\#\vec{P}_i = n_i$, $i \in [n]$, we have that*

$$\boldsymbol{Y}^{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\beta} \simeq \boldsymbol{\alpha}} \binom{|\boldsymbol{\alpha}|}{\boldsymbol{\beta}} \boldsymbol{\theta}_{\mathcal{G}}^{\boldsymbol{\beta}}, \tag{5.5.12}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_i^{\mathrm{T}})_{i\in[n]}^{\mathrm{T}}$, $\boldsymbol{\beta}_i = (\beta_{ij})_{j\in[n_i]}^{\mathrm{T}} \in \mathbb{Z}_{\geq 0}^{n_i}$, $\boldsymbol{\theta}_{\mathcal{G}} = \prod_{i\in[n]} \boldsymbol{\theta}_{\mathcal{G}_i}$ *and we write $\boldsymbol{\alpha} \simeq \boldsymbol{\beta}$ if both* $|\boldsymbol{\alpha}| = |\boldsymbol{\beta}|$ *and $\forall\, i \in [n]$, $|\boldsymbol{\beta}_i| = \alpha_i$.*

Proof. The monomial $\boldsymbol{Y}^{\boldsymbol{\alpha}}$ can be written as

$$\boldsymbol{Y}^{\boldsymbol{\alpha}} = \prod_{i\in[n]} Y_i^{\alpha_i} = \prod_{i\in[n]} \left(\sum_{|\boldsymbol{\beta}_i|=\alpha_i} \binom{\alpha_i}{\boldsymbol{\beta}_i} \boldsymbol{\theta}_{\mathcal{G}_i}^{\boldsymbol{\beta}_i} \right) = \sum_{\boldsymbol{\beta} \simeq \boldsymbol{\alpha}} \boldsymbol{\theta}_{\mathcal{G}}^{\boldsymbol{\beta}} \prod_{i\in[n]} \binom{\alpha_i}{\boldsymbol{\beta}_i}.$$

Noting that

$$\prod_{i\in[n]} \binom{\alpha_i}{\boldsymbol{\beta}_i} = \frac{\prod_{i\in[n]} \alpha_i!}{\prod_{i\in[n]} \prod_{j\in[n_i]} \beta_{ij}!} = \binom{|\boldsymbol{\alpha}|}{\boldsymbol{\beta}},$$

the result follows. $\qquad\qquad\square$

From the previous two results we then deduce the following.

Lemma 5.5.11. *Under the assumptions of Propositions 5.5.9 and 5.5.10, the conditional expected utility is algebraic and can be written as*

$$\bar{u}(\boldsymbol{d} \mid \boldsymbol{\theta}) = \sum_{\boldsymbol{0} <_{lex} \boldsymbol{\alpha} \leq_{lex} \boldsymbol{n}} c_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta} \simeq \boldsymbol{\alpha}} \binom{|\boldsymbol{\alpha}|}{\boldsymbol{\beta}} \boldsymbol{\theta}_{\mathcal{G}}^{\boldsymbol{\beta}}. \tag{5.5.13}$$

Proof. This follows by plugging in equation (5.5.12) into equation (5.5.11). $\qquad\square$

Although straightforward, Lemma 5.5.11 makes a significant generalisation to the theory of the computation of moments in decomposable/additive functions of Cowell et al. [1999] and Nilsson [2001] to multilinear functions of BNs defined as a linear SEM. This result is also connected to the propagation algorithms first developed in Lauritzen [1992]. In that paper, the computation of the first two moments of a much larger class of models (chain graphs with both discrete and continuous variables) is considered. Here, focusing only on a certain class of continuous DAG models, we are able to explicitly compute, through algebraic substitution, any moment of the distribution associated to the graph.

$$((Y_2), (Y_2), (Y_4), (Y_4))$$

$$((Y_1, (Y_1, Y_2)), (Y_2), (Y_4), (Y_4))$$

$$((Y_1, (Y_1, Y_2)), (Y_1, (Y_1, Y_2)), (Y_4), (Y_4))$$

$$((Y_2), (Y_2), (Y_1, (Y_1, Y_4)), (Y_4))$$

$$((Y_1, (Y_1, Y_2)), (Y_2), (Y_1, (Y_1, Y_4)), (Y_4))$$

$$((Y_1, (Y_1, Y_2)), (Y_1, (Y_1, Y_2)), (Y_1, (Y_1, Y_4)), (Y_4))$$

$$((Y_2), (Y_2), (Y_1, (Y_1, Y_4)), (Y_1, (Y_1, Y_4)))$$

$$((Y_1, (Y_1, Y_2)), (Y_2), (Y_1, (Y_1, Y_4)), (Y_1, (Y_1, Y_4)))$$

$$((Y_1, (Y_1, Y_2)), (Y_1, (Y_1, Y_2)), (Y_1, (Y_1, Y_4)), (Y_1, (Y_1, Y_4)))$$

TABLE 5.1: Tuples of dimension 4 with two paths ending in Y_2 and two ending in Y_4 in the graph of Figure 2.1.

Using again the properties of Multinomial coefficients, we can relate equation (5.5.13) to the topology of the graph and its rooted paths. For an $\boldsymbol{\alpha} \in \mathbb{Z}_{\geq 0}^n$, let $\vec{P}^{\boldsymbol{\alpha}}$ be the set of unordered $|\boldsymbol{\alpha}|$-tuples of paths, where in each tuple there are α_i paths ending in Y_i. For each element $P \in \vec{P}^{\boldsymbol{\alpha}}$, let n_P be the sum of n_{P_i}, the number of distinct permutations of paths ending in Y_i in $\vec{P}^{\boldsymbol{\alpha}}$. Then we have that

$$\sum_{\boldsymbol{\beta} \simeq \boldsymbol{\alpha}} \binom{|\boldsymbol{\alpha}|}{\boldsymbol{\beta}} \boldsymbol{\theta}_{\mathcal{G}}^{\boldsymbol{\beta}} = \sum_{P \in \vec{P}^{\boldsymbol{\alpha}}} n_P \prod_{p \in P} \boldsymbol{\theta}_p.$$

Example 5.5.12. 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 summarised in Table 5.1. The associated conditional expectation can be written as the following polynomials, where the i-th monomial corresponds to the tuple in the i-th row of Table 5.1:

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

Note for example that $\theta_{12}\theta_{02}'\theta_{04}'^2$ 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.

Theorem 5.5.13. *Suppose $\boldsymbol{R} = \boldsymbol{Y}$ and that the structural consensus of an IDSS includes a linear SEM over a DAG with vertex set $\{Y_i : i \in [n]\}$, where panel G_i oversees*

the vertex Y_i of $\mathcal{G}$, $i \in [n]$. Suppose the utility consensus consists of a panel separable utility and panel G_i agreed to model its marginal utility with a polynomial utility function of degree $n_i \in \mathbb{Z}_{>0}$, $i \in [n]$. Let $l_i = \#\vec{P}_i$ and $l_{\mathcal{G}} = \sum_{i \in [n]} l_i$, $i \in [n]$. Let $\boldsymbol{\alpha}_{\mathcal{G}} = (\boldsymbol{\alpha}_i^{\mathrm{T}})^{\mathrm{T}}_{i \in [n]} \in \mathbb{Z}^{l_{\mathcal{G}}}_{\geq 0}$, $\boldsymbol{\alpha}_i = (\alpha_{iP})^{\mathrm{T}}_{P \in \vec{P}_i}$ and $\boldsymbol{n} = (n_i)^{\mathrm{T}}_{i \in [n]} \in \mathbb{Z}^n_{\geq 0}$. If, for every $\boldsymbol{\alpha} \simeq \boldsymbol{n}$ $\boldsymbol{\theta}_{\mathcal{G}} = (\boldsymbol{\theta}_i)^{\mathrm{T}}_{i \in [n]}$, with $\boldsymbol{\theta}_i = (\boldsymbol{\theta}_P)^{\mathrm{T}}_{P \in \vec{P}_i}$, entertains moment independence of order $\boldsymbol{\alpha}$, then the IDSS is score separable.

Proof. Under the conditions of the theorem, the conditional expected utility function can be written as in (5.5.13). The linearity of the expectation operator than implies that

$$\mathbb{E}(\bar{u}(\boldsymbol{d} \mid \boldsymbol{\theta})) = \sum_{\boldsymbol{0} <_{lex} \boldsymbol{\alpha} \leq_{lex} \boldsymbol{n}} c_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta} \simeq \boldsymbol{\alpha}} \binom{|\boldsymbol{\alpha}|}{\boldsymbol{\beta}} \mathbb{E}\left(\boldsymbol{\theta}_{\mathcal{G}}^{\boldsymbol{\alpha}} \mid \boldsymbol{d}\right).$$

Now note that

$$\boldsymbol{\theta}_{\mathcal{G}} = \prod_{i \in [n]} \boldsymbol{\theta}_{\mathcal{G}_i} = \prod_{i \in [n]} \prod_{P \in \vec{P}_i} \boldsymbol{\theta}_P = \prod_{i \in [n]} \prod_{P \in \vec{P}_i} \prod_{Y_l \in P} \theta'_{0l} \prod_{(Y_j, Y_k) \in P} \theta_{jk}.$$

Taking the expectation of $\boldsymbol{\theta}_{\mathcal{G}}^{\boldsymbol{\alpha}}$ we then have that

$$
\begin{aligned}
\mathbb{E}\left(\boldsymbol{\theta}_{\mathcal{G}}^{\boldsymbol{\alpha}} \mid \boldsymbol{d}\right) &= \mathbb{E}\left(\prod_{i \in [n]} \boldsymbol{\theta}_{\mathcal{G}_i}^{\boldsymbol{\alpha}_i} \mid \boldsymbol{d}\right) = \mathbb{E}\left(\prod_{i \in [n]} \prod_{P \in \vec{P}_i} \boldsymbol{\theta}_P^{\alpha_{iP}} \mid \boldsymbol{d}\right) \\
&= \mathbb{E}\left(\prod_{i \in [n]} \prod_{P \in \vec{P}_i} \prod_{Y_l \in P} \theta'^{\alpha_{iP}}_{0l} \prod_{(Y_j, Y_k) \in P} \theta_{jk}^{\alpha_{iP}} \mid \boldsymbol{d}\right), \\
&= \prod_{i \in [n]} \prod_{P \in \vec{P}_i} \prod_{Y_l \in P} \mathbb{E}\left((\theta'_{0l}\theta_{lCh_l})^{\alpha_{iP}} \mid \boldsymbol{d}\right) \prod_{(Y_j, Y_k) \in P \setminus (Y_l, Y_{Ch_l})} \mathbb{E}\left(\theta_{jk}^{\alpha_{iP}} \mid \boldsymbol{d}\right),
\end{aligned}
$$

where the last equality follows from moment independence. Score separability then follows. □

This theorem generalises Theorem 5.5.8 to multilinear utility factorisations and thus to a much larger class of IDSSs.

5.5.3 Conflict Models

In this section we exploit the algebraic structure of the conditional expected utility of some simple IDSSs to deduce which panels' belief specification leads to either agreement or disagreement between the members of the collective.

5.5.3.1 A First Simple Example. Suppose an IDSS simply consists of two random variables, Y_1 and Y_2, assumed independent of each other and overseen by two different

panels. For instance Y_1 might represent the costs deriving from a nuclear accident, whilst Y_2 is the number of people not exposed by the contamination. Suppose the optimal number $d \in \mathbb{Z}_{0 < d \leq D}$, $D \in \mathbb{Z}$, of inhabitants to be evacuated from a nearby village needs to be found. Assume $Y_i \mid \theta_i, d \sim \text{Exp}(\theta_i/d)$, $\theta_i \in \mathbb{R}_{>0}$, $i \in [2]$, so that Y_i follows an exponential distribution with density $\theta_i e^{-\theta_i y_i}$, and let

$$u(y_1, y_2) = k_1 u_1(y_1) + k_2 u_2(y_2),$$

with $u_1(y_1) = -\alpha_1 y_1^2 - \beta_1 y_1 + \gamma_1$, $u_2(y_2) = \alpha_2 y_2^2 + \beta_2 y_2$ and $\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1 \in \mathbb{R}_{\geq 0}$. Recall that $\mathbb{E}(Y_i \mid d, \theta_i) = d/\theta_i$ and $\mathbb{V}(Y_i \mid \theta_i, d) = d^2/\theta_i^2$. Then as the number of people evacuated increases, both the costs and the number of non exposed people increase as well. Furthermore note that u_1 and u_2 are respectively an increasing and decreasing function of their arguments in $\mathbb{R}_{\geq 0}$.

It can be easily deduced that the conditional expected utility for an IDSS so defined is equal to

$$
\begin{aligned}
\bar{u}(d \mid \boldsymbol{\theta}) = \mathbb{E}(u(Y_1, Y_2) \mid \boldsymbol{\theta}, \boldsymbol{d}) &= k_1 \left(-2\frac{\alpha_1 d^2}{\theta_1^2} - \frac{\beta_1 d}{\theta_1} + \gamma_1 \right) + k_2 \left(2\frac{\alpha_2 d^2}{\theta_2^2} + \frac{\beta_2 d}{\theta_2} \right) \\
&= 2\left(\frac{k_2 \alpha_2}{\theta_2^2} - \frac{k_1 \alpha_1}{\theta_1^2} \right) d^2 + \left(\frac{k_2 \beta_2}{\theta_2} - \frac{k_1 \beta_1}{\theta_1} \right) d + k_1 \gamma_1.
\end{aligned}
$$

This is a quadratic function of the decision d. Note that if

$$\frac{k_2 \alpha_2}{\theta_2^2} > \frac{k_1 \alpha_1}{\theta_1^2},$$

then $\bar{u}(d \mid \boldsymbol{\theta})$ is concave. Assuming $k_2 \beta_2/\theta_2 < k_1 \beta_1/\theta_1$, otherwise the maximum of the function tends to zero, $\bar{u}(d \mid \boldsymbol{\theta})$ is maximal for some point d^* within the interval $(0, D]$. Therefore optimal policies will in general consist of a balancing of the two attributes and both panels are likely to agree on such a course of action.

On the other hand when

$$\frac{k_2 \alpha_2}{\theta_2^2} < \frac{k_1 \alpha_1}{\theta_1^2},$$

then $\bar{u}(d \mid \boldsymbol{\theta})$ is convex. In this case then the maximum will either be at D or approaching zero, corresponding to only optimising the cost or the exposition attributes respectively. In this situation it will be difficult for panels to agree on the outputs of the IDSS and any policy choice might be contentious. Note that when the coefficient of d^2 is close to zero, then small changes in the elicitation of the parameters might lead to a change from one of the two cases described above to the other.

5.5.3.2 A More Composite Scenario. Although the previous example shed light on the information that the conditional expected utility function carries about the behaviour of its optimal, the study of this behaviour was based on simple analytic concepts. We now consider a more difficult situation and use results from catastrophe theory [Poston and Stewart, 2014, Zeeman, 1979] in the context of expected utility maximisation similarly to Dodd and Smith [2012] and Smith and Dodd [2012].

Suppose an IDSS is defined as in Section 5.5.3.1 above, with the only difference that the utility consensus now includes the utility factorisation

$$u(y_1, y_2) = u_1(y_1)u_2(y_2).$$

The conditional expected utility can then be computed as

$$
\begin{aligned}
\bar{u}(d \mid \boldsymbol{\theta}) &= \left(-2\frac{\alpha_1 d^2}{\theta_1^2} - \frac{\beta_1 d}{\theta_1} + \gamma_1 \right)\left(2\frac{\alpha_2 d^2}{\theta_2^2} + \frac{\beta_2 d}{\theta_2} \right) \\
&= -4\frac{\alpha_1 \alpha_2}{\theta_1^2 \theta_2^2}d^4 - 2\left(\frac{\alpha_1 \beta_2 \theta_2 + \beta_1 \alpha_2 \theta_1}{\theta_1^2 \theta_2^2} \right)d^3 + \frac{2\gamma_1 \alpha_2 \theta_1^2 - \beta_1 \beta_2 \theta_1 \theta_2}{\theta_2^2 \theta_1^2}d^2 + \frac{\gamma_1 \beta_2}{\theta_2}d.
\end{aligned}
$$

In order to study the behaviour of the maxima of this function we first differentiate this expansion and then equate it to zero. Specifically,

$$\frac{\mathrm{d}}{\mathrm{d}d}\bar{u}(d) = d^3 + \frac{2}{3}\frac{\alpha_2 b_1 + b_2 \alpha_1}{a}d^2 + \frac{b_2 b_1 - 2\alpha_2 c}{2a}d - \frac{b_2 c}{4a},$$

where $a = \alpha_1 \alpha_2$, $b_1 = \beta_1 \lambda_1$, $b_2 = \beta_2 \lambda_2$ and $c = \lambda_1^2 \gamma_1$. By letting $e = \frac{2}{3}\frac{\alpha_2 b_1 + b_2 \alpha_1}{a}$, $f = \frac{b_2 b_1 - 2\alpha_2 c}{2a}$, $g = -\frac{b_2 c}{4a}$ and $z = d + \frac{e}{3}$, the equation above can be set to zero as

$$z^3 + \left(f - \frac{1}{3}e^2 \right)z + \left(g + 2\left(\frac{e}{3}\right)^3 - \frac{ef}{3} \right) = 0.$$

The local maxima of the conditional expected utility can therefore be described, as in Smith and Dodd [2012], by the canonical cusp catastrophe [Zeeman, 1979], where the splitting factor of this cusp catastrophe is $-(f - (1/3)e^2)$. Therefore when

$$f - \frac{1}{3}e^2 \geq 0 \quad \Longleftrightarrow \quad \alpha_1 \left(\frac{19}{54}b_2 b_1 \alpha_2 - b_2^2 \alpha_1 \right) \geq \alpha_2^2 \left(c\alpha_1 + \frac{4}{27}b_1^2 \right),$$

the conditional expected utility has one local maximum only: a compromise between the two attributes.

On the other hand if $f - \frac{1}{3}e^2 < 0$, then for values of $g + 2\left(\frac{e}{3}\right)^3 - \frac{ef}{3}$ close to zero, the maximum of the expected utility function will either tend to $d = 0$ or be $d = D$. In this case the optimal decision will only aim at minimising either the costs or the number of people exposed. In such a situations small changes of the parameters might then lead
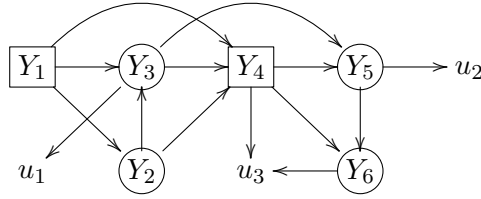
to opposite optimal decision rules and agreement within the collective might be difficult to reach.

5.6 A Note on Some New Symbolic Methods

This section reports recent developments on symbolic methods for inference and decision making in both MIDs and staged trees from Leonelli et al. [2015a] and Görgen et al. [2015], respectively. Although these results will not be discussed within the IDSS framework, we saw in Section 3.6 that both these model classes can be included in the structural consensus of a coherent and distributed IDSS under certain conditions. Therefore the results below can enhance the support an IDSS provides in practice.

5.6.1 Symbolic Evaluation of Influence Diagrams

In this section we introduce a symbolic definition of MIDs and develop a symbolic evaluation algorithm. The implementation of this algorithm with our Maple code is reported in Appendix E.

Let $\mathcal{G}$ be a Multiplicative Influence Diagram (MID) in extensive form with vertex set $\{Y_i, u_j : j \in [m], i \in [n]\}$. Assume each variable Y_i, $i \in [n]$, takes values in $\mathcal{Y}_i = [r_{i-1}]^0$.

5.6.1.1 Polynomial Structure of Expected Utilities. Generalising the work in Darwiche [2003] and Castillo et al. [1995], we introduce a symbolic representation of both the probabilities and the utilities of an MID. For $i \in \mathbb{V}$, $j \in [m]$, and any $y \in \mathcal{Y}_i$, $\pi \in \mathcal{Y}_{\Pi_i}$ and $\sigma \in \mathcal{Y}_{P_j}$, we define the parameters

$$p_{iy\pi} = \mathbb{P}(Y_i = y \mid \mathbf{Y}_{\Pi_i} = \pi), \qquad \psi_{j\sigma} = u_j(\sigma).$$

The first index of $p_{iy\pi}$ and $\psi_{j\sigma}$ refers to the random variable and the utility vertex respectively. The second index of $p_{iy\pi}$ relates to the state of the variable, whilst the third one to the parents' instantiation. The second index of $\psi_{j\sigma}$ corresponds to the instantiation of the arguments of the utility function u_j. We take the indices within π

and σ to be ordered from left to right in decreasing order, so that e.g. p_{6101} for the MID in Figure 2.18, that for convenience we report on top of page 162, corresponds to $\mathbb{P}(Y_6 = 1 \mid Y_5 = 0, Y_4 = 1)$. The *probability* and *utility vectors* are defined as $\boldsymbol{p}_i = (p_{iy\pi})^{\mathrm{T}}_{y \in \mathcal{Y}_i, \pi \in \boldsymbol{\mathcal{Y}}_{\Pi_i}}$ and $\boldsymbol{\psi}_j = (\psi_{j\pi})^{\mathrm{T}}_{\pi \in \boldsymbol{\mathcal{Y}}_{P_j}}$, respectively, $i \in [n]$, $j \in [m]$. Parameters are listed within $\boldsymbol{p}_i$ and $\boldsymbol{\psi}_j$ according to a reverse lexicographic order over their indices (see Appendix D).

Example 5.6.1. The symbolic parametrisation of the MID in Figure 2.18 is summarised in Table 5.2. This is completed by the definition of the criterion weights k_i and h as in equations (2.4.3)-(2.4.4). In Appendix E we report the symbolic definition of this MID using our Maple code.

Because probabilities sum to unity for each i and π, one of the parameters $p_{iy\pi}$ can be written as one minus the sum of the others. Another constraint is induced by equation (2.4.4) on the criterion weights. However in the following, unless otherwise indicated, we suppose that all the parameters are unconstrained. Any unmodelled constraint can be added subsequently when investigating the geometric features of the optimal decision.

Recall that $\bar{u}_i(\boldsymbol{y}_{B_i})$, $i \in [n]$, introduced in Proposition 2.5.10, is the expected utility after either the marginalisation or the maximisation of Y_i and B_i is defined in equation (2.5.1).

Definition 5.6.2. *We define the **conditional expected utility vector** $\bar{\boldsymbol{u}}_i$ as*

$$\bar{\boldsymbol{u}}_i = (\bar{u}_i(\boldsymbol{y}_{B_i}))^{\mathrm{T}}_{\boldsymbol{y}_{B_i} \in \boldsymbol{\mathcal{Y}}_{B_i}}. \tag{5.6.1}$$

In the above parametrisation, $\bar{\boldsymbol{u}}_i$ consists of a vector of polynomials with indeterminates $p_{iy\pi}$, $\psi_{j\sigma}$, k_i and h, for $i \in [n]$, $j \in [m]$, $y \in \mathcal{Y}_i$, $\pi \in \boldsymbol{\mathcal{Y}}_{\Pi_i}$ and $\sigma \in \boldsymbol{\mathcal{Y}}_{P_j}$, whose characteristics are specified in Theorem 5.6.3. Recall that the Decision Sequence (DS) of an MID is the totally ordered sequence over its vertex set and $\mathbb{J} = \{j_i : i \in [m]\}$ is the index set of the vertices preceding a utility node in the DS.

Theorem 5.6.3. *For an MID $\mathcal{G}$ and $i \in [n]$, let $c_i = \prod_{j \in B_i} r_j$, u_l be the first utility node following Y_i in the DS of $\mathcal{G}$ and, for $j \in [m]_{l-1}$, let w_{ij} be the number of random nodes between Y_i and u_j (including Y_i) in the DS of $\mathcal{G}$. Then $\bar{\boldsymbol{u}}_i$ is a vector of dimension c_i whose entries are polynomials including, for $a \in [m]_{l-1}$ and $b \in [a]_{l-1}$, $r_{iba} \in \mathbb{Z}_{\geq 1}$ monomials m_{iba} of degree $d_{iba} \in \mathbb{Z}_{\geq 1}$, where*

$$r_{iba} = \binom{a-l}{b-l} \prod_{j \in [j_a]} r_j, \quad d_{iba} = (b-l) + 2(b-l+1) + w_{ia}, \quad m_{iba} = h^{b-l} m'_{iba}, \tag{5.6.2}$$

with m'_{iba} a square free monomial of degree $2(b-l+1) + w_{ia}$.

$$\boxed{\begin{aligned}
\boldsymbol{p}_2 &= (p_{211}, p_{201}, p_{210}, p_{200})^{\mathrm{T}} \\
\boldsymbol{p}_3 &= (p_{3111}, p_{3011}, p_{3101}, p_{3001}, p_{3110}, p_{3010}, p_{3100}, p_{3000})^{\mathrm{T}} \\
\boldsymbol{p}_5 &= (p_{5111}, p_{5011}, p_{5101}, p_{5001}, p_{5110}, p_{5010}, p_{5100}, p_{5000})^{\mathrm{T}} \\
\boldsymbol{p}_6 &= (p_{6111}, p_{6011}, p_{6101}, p_{6001}, p_{6110}, p_{6010}, p_{6100}, p_{6000})^{\mathrm{T}} \\
\boldsymbol{\psi}_1 &= (\psi_{11}, \psi_{10})^{\mathrm{T}}, \ \boldsymbol{\psi}_2 = (\psi_{21}, \psi_{20})^{\mathrm{T}}, \ \boldsymbol{\psi}_3 = (\psi_{311}, \psi_{301}, \psi_{310}, \psi_{300})^{\mathrm{T}}
\end{aligned}}$$

TABLE 5.2: Parametrisation associated to the diagram in Figure 2.18.

The proof of Theorem 5.6.3 is given in Appendix A.4.3. We say that equation (5.6.2) defines the *structure* of the polynomials of the conditional expected utility $\bar{\boldsymbol{u}}_i$. Specifically, the structure of a polynomial consists of its number of monomials and the number of monomials having a certain degree. In Leonelli et al. [2015a] we demonstrate that the entries of the conditional expected utility vectors of many different MIDs share the same polynomial structure. Two MIDs whose expected utility polynomials have the same structure are said to be *equivalent*. We discuss in Leonelli et al. [2015a] how the notion of equivalent MIDs (equipped with a lattice structure) can be used for instance for model choice and sensitivity analyses.

Since additive utility factorisations can be seen as special cases of multiplicative ones by setting $h = 0$, it follows that the conditional expected utility polynomials of an additive ID are square-free.

Corollary 5.6.4. *In the notation of Theorem 5.6.3, the conditional expected utility $\bar{\boldsymbol{u}}_i$, $i \in [n]$, of an additive ID $\mathcal{G}$ is a vector of dimension c_i whose entries are square-free polynomials of degree $w_{im} + 2$ including, for $a \in [m]_l$, $r_{ia} \in \mathbb{Z}_{\geq 1}$ monomials of degree $w_{ia} + 2$, where $r_{ia} = \prod_{j \in [j_a]} r_j$.*

Proof. This follows directly from Theorem 5.6.3, since an additive factorisation can be derived by setting $n_I - 1$, the exponent of h in equation (2.4.4), equal to zero. This corresponds to fixing $b = l$ in Theorem 5.6.3. $\qquad\square$

Example 5.6.5. For the MID of Figure 2.18 the polynomial structure of the entries of $\bar{\boldsymbol{u}}_5$ can be constructed as follows. Recalling that all the variables are binary, from $B_5 = \{3, 4\}$ it follows that $c_5 = 4$. Thus, $\bar{\boldsymbol{u}}_5$ is a column vector of dimension 4. From $u_2 \equiv u_l$, it follows that

$$r_{522} = 2, \quad r_{523} = 4, \quad r_{533} = 4, \quad d_{522} = 3, \quad d_{523} = 4, \quad d_{533} = 7,$$

where we used the fact that $w_{52} = 1$ and $w_{53} = 2$. All monomials are square-free because the index b of r_{iba} in Theorem 5.6.3 is equal to either l or $l + 1$. Each entry of $\bar{\boldsymbol{u}}_5$ is

a square-free polynomial of degree seven consisting of ten monomials: two of degree 3, four of degree 4 and four of degree 7. The entry $\bar{u}_5(y_3, y_4)$ with $y_3, y_4 \in [1]^0$, of this conditional expected utility can be written as $\bar{u}_5(y_3, y_4) = \bar{u}_5^l(y_3, y_4) + \bar{u}_5^m(y_3, y_4)$ where

$$\bar{u}_5^l = k_2(\psi_{21}p_{51y_4y_3} + \psi_{20}p_{50y_4y_3}) + \sum_{y_5 \in [1]^0} k_3(\psi_{31y_4}p_{61y_5y_4} + \psi_{30y_4}p_{60y_5y_4})p_{5y_5y_4y_3}, \qquad (5.6.3)$$

$$\bar{u}_5^m = hk_2k_3((\psi_{31y_4}p_{610y_4} + \psi_{30y_4}p_{600y_4})\psi_{20}p_{50y_4y_3} + (\psi_{31y_4}p_{611y_4} + \psi_{30y_4}p_{601y_4})\psi_{21}p_{51y_4y_3}). \qquad (5.6.4)$$

An algorithm for computing the polynomials in Theorem 5.6.3 is presented in Section 5.6.1.3.

So far we have assumed that a decision centre has not provided any numerical specification of the uncertainties and the values involved in a given decision problem. This occurs for example if the system is defined through sample distributions of data from different experiments, where probabilities are known with uncertainty. But in practice sometimes the members of the centre are able to elicit the numerical values of some parameters. These can then be simply substituted into the corresponding probability and utility parameters in the system of polynomials constructed in Theorem 5.6.3 employing e.g. a computer algebra system. In such a case the degree of the polynomials and possibly their number of monomials can decrease dramatically. We present in Section 5.6.1.2 a plausible numerical specification of the probabilities associated with the MID in Figure 2.18.

5.6.1.2   A New Algebra For MIDs. Computing the polynomials in Theorem 5.6.3 is a well known NP-hard problem [Cooper, 1990]. Here we develop an algorithm based on three operations which exploit the polynomial structure of expected utilities and use only linear algebra calculus. The Maple code for their implementation is in Appendix E. In contrast to the symbolic algorithms mentioned in Section 5.1, our algorithm sequentially computes only monomials that are part of the MID's expected utilities.

The operations we introduce below entail a change of dimension in the probability, utility and conditional expected utility vectors because some of the components in their subvectors are duplicated. These duplications are called `EUDuplicationPsi` and `EUDuplicationP`. These are required in order to multiply parameters associated to compatible instantiations only. We do not report here the working of these procedure, since these are not fundamental for following developments, but we include their Maple implementations in Appendix E. More details about these can also be found in Leonelli et al. [2015a].

The first of the three operations we introduce is `EUMultiSum`, which computes a weighted multilinear sum between a utility vector and a conditional expected utility. For additive IDs, `EUMultiSum` consists of only summations since $h = 0$. In the algorithm of Section 5.6.1.3, an `EUMultiSum` operation is associated to every utility vertex of the MID.

Definition 5.6.6 (EUMultiSum). *For $i \in [n]$, let $\bar{\boldsymbol{u}}_{i+1}$ be a conditional expected utility vector and $\boldsymbol{\psi}_j$ the utility vector of node u_j, $j \in [m]$, succeeding Y_i in the DS. The EUMultiSum, $+^{\mathrm{EU}}$, between $\bar{\boldsymbol{u}}_{i+1}$ and $\boldsymbol{\psi}_j$ is defined as*

1. *$\bar{\boldsymbol{u}}'_{i+1}, \boldsymbol{\psi}'_j \longleftarrow EUDuplicationPsi(\cdot)$;*

2. *$h \cdot k_j \cdot (\bar{\boldsymbol{u}}'_{i+1} \circ \boldsymbol{\psi}'_j) + k_j \cdot \boldsymbol{\psi}'_j + \bar{\boldsymbol{u}}'_{i+1}$, where $\circ$ and $\cdot$ denote the Schur (or element by element) and the scalar products respectively.*

The second operation, `EUMarginalisation`, is applied to any random vertex of the MID.

Definition 5.6.7 (EUMarginalisation). *For $i \in \mathbb{V}$, let $\bar{\boldsymbol{u}}_{i+1}$ be a conditional expected utility vector and $\boldsymbol{p}_i$ a probability vector. The EUMarginalisation, Σ^{EU}, between $\bar{\boldsymbol{u}}_{i+1}$ and $\boldsymbol{p}_i$ is defined as*

1. *$\bar{\boldsymbol{u}}'_{i+1}, \boldsymbol{p}'_i \longleftarrow EUDuplicationP(\cdot)$;*

2. *$I'_i \times (\bar{\boldsymbol{u}}'_{i+1} \circ \boldsymbol{p}'_i)$, where $\times$ denotes the standard matrix product and I'_i is a matrix with $c_{i+1}s_i/r_i \in \mathbb{Z}_{\geq 1}^2$ rows and $c_{i+1}s_i$ columns defined as*

$$I'_i = \left(\begin{pmatrix} \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{1} & \cdots & \mathbf{0} \end{pmatrix} \cdots \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} \end{pmatrix} \right)^{\mathrm{T}}$$

where $\mathbf{1}$ and $\mathbf{0}$ denote row vectors of dimension r_i with all entries equal to one and zero respectively and $s_i = \prod_{k \in \{\Pi_i \backslash B_{i+1}\}} r_k$.

The last operation is a maximisation of $\bar{\boldsymbol{u}}_{i+1}$ over the space $\mathcal{Y}_i$, $i \in \mathbb{D}$, for any element of $\boldsymbol{\mathcal{Y}}_{\Pi_i}$.

Definition 5.6.8 (EUMaximisation). *For $i \in \mathbb{D}$, let $\bar{\boldsymbol{u}}_{i+1}$ be a conditional expected utility vector. An EUMaximisation over $\mathcal{Y}_i$, $\max_{\mathcal{Y}_i}^{\mathrm{EU}}$, is defined by the following steps:*

1. *set $y_i^*(\pi) = \arg\max_{\mathcal{Y}_i} \bar{\boldsymbol{u}}_{i+1}$, for $\pi \in \boldsymbol{\mathcal{Y}}_{\Pi_i}$;*

---

[2]This number is integer since $c_{i+1} = r_i a_{i+1}$, for an $a_{i+1} \in \mathbb{Z}_{\geq 1}$.

$k_1 = 0.2,$	$h = 2.6,$	$\psi_{311} = 0,$	$p_{5101} = 0.9$	
$k_2 = 0.3,$	$\psi_{20} = 1,$	$\psi_{300} = 1,$	$p_{5110} = 0.2,$	$p_{6100} = 0.3$
$k_3 = 0.5,$	$\psi_{21} = 0,$	$\psi_{310} = 0.4,$	$p_{5100} = 0.6$	

TABLE 5.3: Numerical specification of a subset of the indeterminates associated to the diagram of Figure 2.18.

2. $I_i^* \times \bar{\boldsymbol{u}}_{i+1}$, where I_i^* is a matrix with $c_{i+1}/r_i \in \mathbb{Z}_{\geq 1}$ rows and c_{i+1} columns defined as

$$I_i^* = \left(\left(\begin{array}{cccc} \boldsymbol{e}_{\boldsymbol{y}_i^*(1)} & \boldsymbol{0} & \cdots & \boldsymbol{0} \end{array} \right) \left(\begin{array}{cccc} \boldsymbol{0} & \boldsymbol{e}_{\boldsymbol{y}_i^*(2)} & \cdots & \boldsymbol{0} \end{array} \right) \cdots \left(\begin{array}{ccc} \boldsymbol{0} & \cdots & \boldsymbol{e}_{\boldsymbol{y}_i^*(c_{i+1}/r_i)} \end{array} \right) \right)^{\mathrm{T}}$$

where $\boldsymbol{e}_{\boldsymbol{y}_i^*(\pi)}$, $\pi \in [c_{i+1}/r_i]$, is a row vector of dimension r_i whose entries are all zero but the one in position $y_i^*(\pi)$, which is equal to one.

The first item in Definition 5.6.8 is the critical part of the operation `EUMaximisation`. It is not the scope of this thesis to present a methodology to identify such expected utility maximising decisions. We simply assume that these can be found. However we note that, within our symbolic approach, polynomial optimisation and semi-algebraic methods can be used to determine such optimal decisions [Parrilo, 2000]. Once these are identified, `EUMaximisation` drops the polynomials associated with the non optimal course of action. Alternatively, we can think of the evaluation of the MID as the computation of the expected utility polynomial of a specific policy. This is for example all that can be done whenever decision rules have been committed to a priori. In this case, `EUMaximisation` deletes the decisions that are not adopted within the given policy. The Maple function `EUMaximisation` in Appendix E calls a subfunction `Maximise`, which randomly picks optimal decisions.

Example 5.6.9. To illustrate the working of `EUMaximisation`, suppose for the MID in Figure 2.18 that a decision centre has provided the specifications summarised in Table 5.3 together with the *qualitative* beliefs $p_{5111} = p_{6011}$ and $p_{6010} = p_{6001}$. These in general cannot be implemented in a non-symbolic approach to decision making problems. By plugging in these numeric values and constraints into equations (5.6.3) and (5.6.4), the centre would choose to evacuate ($Y_4 = 1$) for combinations of the parameters leading to the coloured areas of Figure 5.1, assuming $Y_3 = 1$ in Figure 5.1a and $Y_3 = 0$ in Figure 5.1b. The geometric structure of the regions often gives insights about the maximisation process. Assuming $Y_3 = 1$, if the members of the centre believe that $\psi_{301}, p_{5111}, p_{6001} \in [0, 1/2]$, then Figure 5.1a shows that evacuation will be the optimal choice. Conversely, when $Y_3 = 0$, such a range for the indeterminates would not uniquely identify an optimal course of action. We can however envisage the algorithm to work over the two sub-regions of Figure 5.1b separately. The algorithm would then output
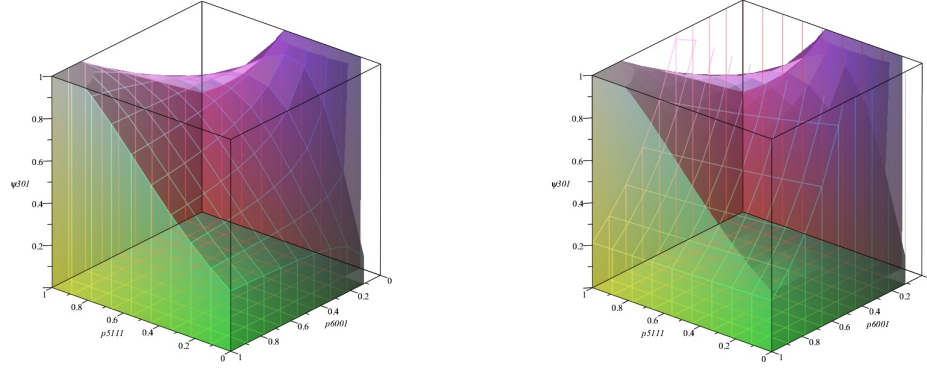
(A) Optimal regions in the case $Y_3 = 1$.   (B) Optimal regions in the case $Y_3 = 0$.

FIGURE 5.1: Regions determining the combinations of parameters leading to an optimal decision of evacuating (coloured regions) and not evacuating (white regions) for the evaluation of the diagram in Figure 2.18 given the partial numeric specification in Table 5.3.

different optimal courses of action for different combinations of the unknown parameters. We plan to develop a systematic methodology to address these issues in later work.

Each of the above three operations changes the conditional expected utility vectors and their entries in a specific way we formalise below.

Proposition 5.6.10. *For $i \in [n]$, let $\bar{\boldsymbol{u}}_{i+1}$ be a conditional expected utility vector whose entries have the polynomial structure of equation (5.6.2) and let u_j be the vertex preceding Y_{i+1} in the DS. Then, in the notation of Theorem 5.6.3,*

- $\max_{\mathcal{Y}_i}^{\mathrm{EU}} \bar{\boldsymbol{u}}_{i+1}$ *has dimension c_{i+1}/r_i and its entries do not change polynomial structure;*

- $\bar{\boldsymbol{u}}_{i+1} +^{\mathrm{EU}} \boldsymbol{\psi}_j$ *has dimension $c_{i+1}s_i^{\mathbb{U}}$, where $s_i^{\mathbb{U}} = \prod_{k \in \{P_j \setminus B_{i+1}\}} r_k$, and each of its entries includes $r_{(i+1)ba}$ monomials of degree $d_{(i+1)ba}$, $r_{(i+1)ba}$ monomials of degree $d_{(i+1)ba} + 3$ and one monomial of degree 2;*

- $\bar{\boldsymbol{u}}_{i+1} \Sigma^{\mathrm{EU}} \boldsymbol{p}_i$ *has dimension $c_{i+1}s_i/r_i$, where $s_i = \prod_{k \in \{\Pi_i \setminus B_{i+1}\}} r_k$, and each of its entries includes $r_i r_{(i+1)ba}$ monomials of degree $d_{(i+1)ba} + 1$.*

This result directly follows from the definition of the above three operations whose effect on the polynomials associated to the diagram in Figure 2.18 is illustrated below.

5.6.1.3 A Fast Evaluation Algorithm. The algorithm for the evaluation of MIDs is given in Algorithm 5.6.1. This receives as input the DS of the MID, S say, the sets $\mathbb{J}$, $\mathbb{V}$ and $\mathbb{D}$, and the vectors $\boldsymbol{p} = (\boldsymbol{p}_i)_{i \in [n]}^{\mathrm{T}}$, $\boldsymbol{\psi} = (\boldsymbol{\psi}_j)_{j \in [n]}^{\mathrm{T}}$ and $\boldsymbol{k} = (h, k_j)_{j \in [m]}^{\mathrm{T}}$. We

---

Algorithm 5.6.1: SYMBOLICEXPECTEDUTILITY$(\mathbb{J}, S, \boldsymbol{p}, \boldsymbol{\psi}, \boldsymbol{k}, \mathbb{V}, \mathbb{D})$

$$\bar{\boldsymbol{u}}_{n+1} = (0) \tag{1}$$

for $k \leftarrow n$ **downto** 1 $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (2)

$\textbf{do}\ \begin{cases} \textbf{for}\ l \leftarrow m\ \textbf{downto}\ 1 & (3) \\ \textbf{do}\ \begin{cases} \textbf{if}\ k = j_l & (4) \\ \quad\textbf{then}\ \begin{cases} \textbf{if}\ k \in \mathbb{D} & (5) \\ \quad\textbf{then}\ \left\{\bar{\boldsymbol{u}}_k = \max_{\mathcal{Y}_k}^{\mathrm{EU}}(\bar{\boldsymbol{u}}_{k+1} +^{\mathrm{EU}}\boldsymbol{\psi}_l\right) & (6) \\ \quad\textbf{else}\ \left\{\bar{\boldsymbol{u}}_k = \boldsymbol{p}_k\ \Sigma^{\mathrm{EU}}(\bar{\boldsymbol{u}}_{k+1} +^{\mathrm{EU}}\boldsymbol{\psi}_l\right) & (7) \end{cases} \\ \textbf{else if}\ k \in \mathbb{D} & (8) \\ \textbf{then}\ \left\{\bar{\boldsymbol{u}}_k = \max_{\mathcal{Y}_k}^{\mathrm{EU}}\bar{\boldsymbol{u}}_{k+1}\right. & (9) \\ \textbf{else}\ \left\{\bar{\boldsymbol{u}}_k = \boldsymbol{p}_k\ \Sigma_{\mathcal{Y}_k}^{\mathrm{EU}}\bar{\boldsymbol{u}}_{k+1}\right. & (10) \end{cases} \end{cases}$

return $(\bar{\boldsymbol{U}}_1)$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (11)

---

implemented this algorithm in the `SymbolicExpectedUtility` function in Appendix E.[3]. This corresponds to a symbolic version of the backward induction procedure over the elements of the DS explicated in Proposition 2.5.10. At each inductive step, a utility vertex is considered together with the variable that precedes it in the DS.

In line (1) the conditional expected utility $\bar{\boldsymbol{u}}_{n+1}$ is initialised to (0). Lines (2) and (3) index a reverse loop over the indices of both the variables and the utility vertices respectively (starting from n and m). If the current index corresponds to a variable preceding a utility vertex in the DS (line 4), then the algorithm jumps to lines (5)-(7). Otherwise it jumps to lines (8)-(10). In the former case, the algorithm computes, depending on whether or not the variable is controlled (line 5), either an `EUMaximisation` over $\mathcal{Y}_k$ (line 6) or an `EUMarginalisation` (line 7) with $\boldsymbol{p}_k$, jointly to an `EUMultiSum` with $\boldsymbol{\psi}_l$. In the other case, `EUMaximisation` and `EUMarginalisation` operations are performed without `EUMultiSum`.

Example 5.6.11. For the MID in Figure 2.18 the `SymbolicExpectedUtility` function first considers the random vertex Y_6 which precedes the utility vertex U_3 and therefore first calls the `EUMultiSum` function. `EUDuplicationPsi` in this case entails that $\bar{u}_7$ is duplicated four times and

$$\bar{\boldsymbol{U}}_7 +^{\mathrm{EU}}\boldsymbol{\psi}_3 = \begin{pmatrix} k_3\psi_{11} & k_3\psi_{01} & k_3\psi_{10} & k_3\psi_{00} \end{pmatrix}^{\mathrm{T}}. \tag{5.6.5}$$

---

[3]Note that the inputs of these functions in Maple are different. The ones chosen here are able to more concisely report their working.

Then, the rhs of Equation (5.6.5) needs to be duplicated via `EUDuplicationP` and

$$\bar{\boldsymbol{u}}_6^{\mathrm{T}} = I_6' \times \bar{\boldsymbol{u}}_6^t \circ \boldsymbol{p}_6 = (k_3\psi_{31j}p_{61ij} + k_3\psi_{30j}p_{60ij})_{i,j\in[1]^0}^{\mathrm{T}},$$

where $\bar{\boldsymbol{u}}_6^t$ is equal to the duplicated version of the rhs of equation (5.6.5). The vector $\bar{\boldsymbol{u}}_6$ has dimension four and its entries include two monomials of degree 3. Since the random vertex Y_5 is the unique parent of U_2 the `SymbolicExpectedUtility` function follows the same steps as before. `EUMultiSum` is first called which gives as output

$$\bar{\boldsymbol{u}}_6 +^{\mathrm{EU}} \boldsymbol{\psi}_2 = h \cdot k_2 \cdot \bar{\boldsymbol{u}}_6 \circ \left(\begin{array}{cccc} \psi_{21} & \psi_{20} & \psi_{21} & \psi_{20} \end{array} \right)^{\mathrm{T}} + \bar{\boldsymbol{u}}_6 + k_2 \cdot \left(\begin{array}{cccc} \psi_{21} & \psi_{20} & \psi_{21} & \psi_{20} \end{array} \right)^{\mathrm{T}}$$

The polynomial $\bar{\boldsymbol{u}}_6 +^{\mathrm{EU}} \boldsymbol{\psi}_2$ is the sum of two monomials of degree 3 inherited from $\bar{\boldsymbol{u}}_6$, of two monomials of degree 6 (from the first term on the rhs of 5.6.11) and one monomial of degree 2 (from the last term on the rhs of 5.6.11). Its dimension is equal to four (i.e. no `EUDuplicationPsi` is required). `EUMultiSum` manipulates the conditional expected utility vector according to Proposition 5.6.10. Then, the `EUMarginalisation` function computes

$$\bar{\boldsymbol{u}}_5 = I_5' \times \left(\left(\begin{array}{cc} \bar{\boldsymbol{u}}_6 +^{\mathrm{EU}} \boldsymbol{\psi}_2 & \bar{\boldsymbol{u}}_6 +^{\mathrm{EU}} \boldsymbol{\psi}_2 \end{array} \right)^{\mathrm{T}} \circ \boldsymbol{p}_5 \right).$$

Each entry of $\bar{\boldsymbol{u}}_5$ has twice the number of monomials of the entries of $\bar{\boldsymbol{u}}_6 +^{\mathrm{EU}} \boldsymbol{\psi}_2$ and each monomial of $\bar{\boldsymbol{u}}_5$ has degree $d+1$, where d is the degree of each monomial of $\bar{\boldsymbol{u}}_6 +^{\mathrm{EU}} \boldsymbol{\psi}_2$ (whose entries are homogeneous polynomials). These vectors also have the same dimension. Thus, this `EUMarginalisation` changes the conditional expected utility vector according to Proposition 5.6.10. The polynomial in a generic entry of $\bar{\boldsymbol{u}}_5$ was shown in equations (5.6.3)-(5.6.4).

The algorithm then considers the controlled variable Y_4. Since $4 \notin \mathbb{J}$, Y_4 is not the argument of a utility function with the highest index and therefore the algorithm calls the `EUMaximisation` function. Suppose the optimal decisions are identified to be $Y_4 = 1$ when $Y_3 = 1$ and $Y_4 = 0$ when $Y_3 = 0$. The evaluation would then suggest that the population is evacuated whenever a high level of deposition is observed and that people are not evacuated if the deposition is low. Then `EUMaximisation` would return $\bar{\boldsymbol{u}}_4 = I_4^* \times \bar{\boldsymbol{u}}_5$, where I_4^* is a 2×4 matrix with ones in positions $(1,1)$ and $(2,4)$ and zeros otherwise. Proposition 5.6.10 is respected since the entries of $\bar{\boldsymbol{u}}_4$ have the same polynomial structure of those of $\bar{\boldsymbol{u}}_5$ and $\bar{\boldsymbol{u}}_4$ has dimension 2.

The `SymbolicExpectedUtility` function then applies in sequence the operations defined in Section 5.6.1.2. For the MID in Figure 2.18 it sequentially computes the following

quantities:

$$\bar{\boldsymbol{u}}_3^t = h \cdot k_1 \cdot \bar{\boldsymbol{u}}_4 \circ \boldsymbol{\psi}_1 + \bar{\boldsymbol{u}}_4 + k_1 \cdot \boldsymbol{\psi}_1, \quad \bar{\boldsymbol{u}}_3 = I_3' \times \left(\left(\begin{array}{cccc} \bar{\boldsymbol{u}}_3^t & \bar{\boldsymbol{u}}_3^t & \bar{\boldsymbol{u}}_3^t & \bar{\boldsymbol{u}}_3^t \end{array} \right)^{\mathrm{T}} \circ \boldsymbol{p}_3 \right),$$

$$\bar{\boldsymbol{u}}_2 = I_2' \times (\bar{\boldsymbol{u}}_3 \circ \boldsymbol{p}_2), \qquad\qquad \bar{\boldsymbol{u}}_1 = \left(\begin{array}{cc} 1 & 0 \end{array} \right) \times \bar{\boldsymbol{u}}_2,$$

assuming the optimal initial decision is $Y_1 = 1$.

Interestingly, using the new algebra we introduced in Section 5.6.1.2, the evaluation of an MID can be written as a simple algebraic expression. For example, the evaluation of the MID in Figure 2.18 can be written as

$$\bar{\boldsymbol{U}}_1 = \max_{\mathcal{Y}_1}^{\mathrm{EU}} \left(\boldsymbol{p}_2 \; \Sigma^{\mathrm{EU}} \left(\boldsymbol{p}_3 \; \Sigma^{\mathrm{EU}} \left(\boldsymbol{\psi}_1 +^{\mathrm{EU}} \max_{\mathcal{Y}_4}^{\mathrm{EU}} \left(\boldsymbol{p}_5 \; \Sigma^{\mathrm{EU}} (\boldsymbol{\psi}_2 +^{\mathrm{EU}} (\boldsymbol{p}_6 \; \Sigma^{\mathrm{EU}} \boldsymbol{\psi}_3))) \right) \right) \right) \right)$$

and this polynomial can be evaluated with `SymbolicExpectedUtility`.

Although we have developed in this section a symbolic approach for the evaluation of extensive form MIDs, in Leonelli et al. [2015a] we formalise both the symbolic and the polynomial interpretation of the manipulations (edge reversal and barren node removal) of non extensive form diagrams. In that paper we further deduce the polynomial and symbolic interpretation of the sufficiency theorem of Smith [1989a], that can be used to greatly simplify the evaluation of an MID by deleting some of its vertices.

5.6.1.4 Asymmetric Decision Problems. In Section 2.5.2 we claimed that most of the graphical representations of asymmetric decision problems lose the transparency associated to IDs. In this section we start to demonstrate one of the great advantages of symbolic approaches, namely that the complexity of both the polynomial structure and the symbolic definition of an asymmetric problem are simpler than the ones using a symmetric embedding of the problem. We discuss this feature in greater length in Section 5.6.2 below.

Here we characterise an asymmetry between two chance nodes which, depending on the stage of the evaluation, may entail setting equal to zero monomials in either some or all rows of the conditional expected utility vector. This therefore implies that the polynomial structure of the conditional expected utility vectors and at times also their dimension change. We present the result for elementary asymmetries of the following form: if $Y_i = y_i$ then $Y_j \neq y_j$, $i \neq j$ (see Section 2.5.2 for an example of such asymmetries). Composite asymmetries are unions of simple asymmetries and the features of the conditional expected utility vectors in more general cases can be deduced through a sequential application of Theorem 5.6.12.

Theorem 5.6.12. *Let $\mathcal{G}$ be an MID, Y_i and Y_j be two random variables with $j > i$, U_x be the first utility node that follows Y_j in the DS. Assume the asymmetry $Y_i = y_i \Rightarrow Y_j \neq y_j$ holds and that k and z are the highest indices such that $j \in B_k$ and $i \in B_z$ and assume $k > j$. Then*

- *for $t \in [z]_j$, each row of $\bar{u}_t$ has $\prod_{s \in B_t \setminus \{i \cup j\}} r_s$ rows with no monomials;*

- *for $t \in [j]_i$, each row of $\bar{u}_t$ has $\prod_{s \in B_t \setminus \{i\}} r_s$ rows with polynomials of a different structure. Specifically, these consists of, in the notation of Theorem 5.6.3, r'_{tba} monomials of degree d_{tba}, where for $a \in [m]_{x-1}$ and $b \in [a]_{l-1}$*

$$r'_{tba} = \left(\binom{a-x}{b-l} - 1 \right) \prod_{s \in [j_a]_{t-1}} r_s / r_j;$$

- *for $t \in [i]$, each row of $\bar{u}_t$ has in the notation of Theorem 5.6.3 r''_{tba} monomials of degree d_{tba}, where for $a \in [m]_{x-1}$ and $b \in [a]_{l-1}$*

$$r''_{tba} = \left(\binom{a-x}{b-l} - 1 \right) \prod_{s \in [j_a]_{t-1}} r_s / (r_j \cdot r_i).$$

The proof of this theorem is provided in Appendix A.4.4. Corollary 5.6.13 gives a characterisation of simple asymmetries between any two variables, whether they are controlled or non-controlled. This follows from Theorem 5.6.12 since controlled variables can be thought of as a special case of random ones.

Corollary 5.6.13. *In the notation of Theorem 5.6.3 and under the assumptions of Theorem 5.6.12, with the difference that Y_i and Y_j are two variables, controlled or non-controlled, we have that*

- *for $t \in [z]_j$, each row of $\bar{u}_t$ has $\prod_{s \in B_t \setminus \{i \cup j\}} r_s$ rows with no monomials;*

- *for $t \in [j]_i$, $\bar{u}_t$ has at most $\prod_{s \in B_t \setminus \{i\}} r_s$ rows with polynomials of a different structure. Specifically, these consists of between r'_{tba} and r_{tba} monomials of degree d_{tba}, for $a \in [m]_{x-1}$ and $b \in [a]_{l-1}$;*

- *for $t \in [i]$, some rows of $\bar{u}_t$ have a number of monomials of degree d_{tba} between r''_{tba} and r_{tba}, for $a \in [m]_{x-1}$ and $b \in [a]_{l-1}$.*

Example 5.6.14 (Example 2.5.11 continued)**.** Consider the asymmetric problem described in Figure 2.19. The symbolic imposition of the asymmetric constraints reduces from ten to three the number of monomials in $\bar{u}_5$ which becomes

$$k_3 \psi_{311} p_{6111} p_{511i} + k_2 \psi_{21} p_{511i} + h k_2 k_3 \psi_{311} \psi_{21} p_{6111} p_{511i}, \quad i \in [1]^0.$$

Suppose the `EUMaximisation` suggested that $Y_4 = 0$ is optimal if $Y_3 = 1$ and that $Y_4 = 1$ is preferred if $Y_3 = 0$. The entry of $\bar{u}_3$ for which $Y_2 = 1$ and $Y_1 = 1$ can be written as

$$\sum_{i, \in [1]^0} ((k_2 \psi_{21} + k_3 \psi_{311} p_{6111} (1 + h k_2 \psi_{21})) p_{5110} p_{3011} + k_1 \psi_{1i} p_{3i11}$$

$$+ h k_1 k_3 \psi_{11} p_{5101} p_{3111} ((1 + k_2 \psi_{21}) \psi_{3j0} p_{6j10})).$$

This polynomial consists of only nine monomials. This, compared with the number of monomials in the symmetric case, 42, means that even in this small problem we obtain a reduction of the number of monomials by over three quarters.

So the example above illustrates that under asymmetries the polynomial representation is simpler than standard methods but still able to inform decision centres about the necessary parameters to elicit. Finally, it is possible to develop a variant of Algorithm 5.6.1 which explicitly takes into account the asymmetries of the problem *during* its evaluation. Note that this approach would be computationally even more efficient, since it would require the computation of a smaller number of monomials/polynomials.

5.6.2 A Differential Approach to Inference in Staged Trees

5.6.2.1 Symbolic Definition of Staged Trees. As discussed in Section 5.1, symbolic methods have proved useful in the study of BNs, despite coming with a considerable computational cost. In this section we develop a symbolic approach for *staged trees* [see e.g. Smith and Anderson, 2008, and Section 2.3.2.6] and we demonstrate that such difficulties are eased, because of the more intuitive parametrisation associated to these models. In addition, because of the wide variety of possible hypotheses they embody, staged trees are necessarily models over much smaller state spaces than BNs. Since the size of the model space is the main computational issue for symbolic approaches associated with BNs, it follows that trees can be very practical contexts to investigate symbolic inferential queries.

Borrowing ideas from Pistone et al. [2001], we define the *interpolating polynomial* of a staged tree.

Definition 5.6.15. *Let* $\mathcal{T} = (V, E)$ *be a staged tree with primitive probabilities* θ_e, $e \in E$, *and set of root-to-leaf paths* $\Lambda(s_0, \mathcal{T})$. *We call* $\Lambda(e) = \{r \in \Lambda(s_0, \mathcal{T}) \mid e \in E(r)\}$ *an* edge centred *event, and set* $\lambda_e(r)$, *for* $e \in E$, *to be an indicator of* $r \in \Lambda(e)$. *We call*

$$c_{\mathcal{T}}(\theta, \lambda) = \sum_{r \in R_{\mathcal{T}}} \prod_{e \in E(r)} \lambda_e(r) \theta_e$$

the **interpolating polynomial** of $\mathcal{T}$.

The interpolating polynomial is a sum of probabilities of atomic events with indicators for certain conditional events happening or not happening. Even though all these unknowns sum to one, in our symbolic approach we treat them just like indeterminates. Note that the indicators $\lambda_e(r)$ on the edges $e \in E(r)$ are associated to the (conditional) event represented by e, having probability θ_e. This notation is apparently redundant, but turns out to be useful below. We observe that this redundancy is one of the great advantages of a staged tree: whilst Darwiche [2003] needs to compute conditional probabilities of all *compatible parent structures* of an event, which is a rather obscure concept in a symbolic framework, and Castillo et al. [1996] computes the product space of any indeterminates' combination regardless of their meaning, a tree visualisation of our model gives us the necessary structure immediately: events can be simply read from the paths in the graph.

Example 5.6.16. The interpolating polynomial of the tree in Figure 2.7 (on page 150), not considering the underlying stage structure, is equal to (for ease of notation we omit the indicator functions)

$$c'_{\mathcal{T}}(\theta) = \theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{11}\theta_{30} + \theta_{01}\theta_{10}\theta_{41} + \theta_{01}\theta_{10}\theta_{40} + \theta_{00}\theta_{21} + \theta_{00}\theta_{20}. \qquad (5.6.6)$$

When plugging in the conditional independence constraints, we obtain the interpolating polynomial of the staged tree as

$$c_{\mathcal{T}}(\theta) = \theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{11}\theta_{30} + \theta_{01}\theta_{10}\theta_{31} + \theta_{01}\theta_{10}\theta_{30} + \theta_{00}\theta_{11} + \theta_{00}\theta_{10}, \qquad (5.6.7)$$

$$= \theta_{01}(\theta_{11}(\theta_{31} + \theta_{30}) + \theta_{10}(\theta_{31} + \theta_{30})) + \theta_{00}(\theta_{11} + \theta_{10}). \qquad (5.6.8)$$

Conversely, the interpolating polynomial of the BN representation of the problem described by this staged tree, ignoring the conditional independence structure, equals

$$\begin{aligned} c_{\mathrm{BN}}(\theta) =& \theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{11}\theta_{30} + \theta_{01}\theta_{10}\theta_{41} + \theta_{01}\theta_{10}\theta_{40} \\ & + \theta_{00}\theta_{21}\theta_{51} + \theta_{00}\theta_{21}\theta_{50} + \theta_{00}\theta_{20}\theta_{61} + \theta_{00}\theta_{20}\theta_{60}, \end{aligned} \qquad (5.6.9)$$

We now look at interpolating polynomials from an algebraic viewpoint. Because of the product space structure of the sample space of a BN, the interpolating polynomial of a BN with n vertices equals the sum of monomials each of which is of degree $2n$ and so *homogeneous*. Moreover, the stage structure of a tree associated to a BN model as in Figure 2.8 is such that no two vertices along the same directed path are in the same stage, in fact stages exist only along *orthogonal cuts* [Thwaites and Smith, 2015]. Thus, the interpolating polynomial of a BN is also square-free. This is shown for instance by the interpolating polynomial c_{BN} in equation (5.6.9), an homogeneous polynomial of

degree 3 whose number of terms equals the number of paths in its tree representation. Importantly this polynomial has been simply read from the event tree by first multiplying over all primitive probabilities along one root-to-leaf path, and then summing over all of these paths. This is a lot easier done using Figure 2.8 (on page 36) than in the associate BN representation, where we would have had to sum over compatible parent configurations.

Conversely, interpolating polynomials of staged trees are not necessarily homogeneous and square-free. For instance, the interpolating polynomial in equation (5.6.7) is now inhomogeneous with total degree 3 and monomial terms having degree 2 and 3. Although our staged tree examples have a square-free structure, non multilinear interpolating polynomials arise from a tree where two vertices in the same root-to-leaf path are in the same stage [as shown in Görgen et al., 2015]. For the purpose of this section we focus only on staged trees whose interpolating polynomial is square-free.

Notice that $c_\mathcal{T}$ can be easily factorised in equation (5.6.8) by simply following the structure of the underlying graph. In Brandherm and Jameson [2004] polynomials of this type are called *factored*. This representation entails great computational advantages since the compilation into an AC is almost instantaneous. Whilst for BNs the factored representation might be difficult to obtain, it comes almost for free in tree models.

We observe that the graphical simplicity of staged tree models in comparison to an uncoloured tree or trees associated to BNs is also reflected algebraically: the polynomial in (5.6.7) has fewer indeterminates than the one in (5.6.6) and a lot fewer than the polynomial associated to a tree which is derived from a BN in (5.6.9). This is because in the BN the redundancy of atoms gives rise to redundant terms.

The definition of interpolating polynomials enables us to compute the probability of any event associated to a tree.

Lemma 5.6.17. *For any event A represented by a set of root-to-leaf paths R_A in a staged tree $\mathcal{T}$,*

$$\mathbb{P}(A) = \sum_{r \in R_A} \prod_{e \in E(r)} \lambda_e(r)\theta_e = c_\mathcal{T}(\theta, \lambda|_{R_A}),$$

where $\lambda|_{R_A}$ indicates that $\lambda_e(r) = 1$ for all $e \in E(r)$ with $r \in R_A$, and else zero.

This result holds since the probability of an event is equal to the sum of the probabilities of the associated root-to-leaf paths.

Example 5.6.18. Suppose we are interested in calculating the probability of political disruption. This is captured by the set $R_A = \Lambda(e_{31}) \cup \Lambda(e_{41})$ corresponding to all root-to-leaf paths going through an edge labelled 'yes', which translates in summing all terms

in (5.6.7) including θ_{31}. Therefore $\mathbb{P}(A) = \theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{12}\theta_{31}$, again omitting the λ indicators.

5.6.2.2 The Differential Approach. We are now able to provide a probabilistic semantic, just as Darwiche [2003] for BNs, to the derivatives of polynomials associated to staged trees. For ease of notation we let in this section $\lambda_e = \lambda_e(r)$.

Proposition 5.6.19. *For equally coloured edges $e \in E$ and an event A represented by the root-to-leaf paths R_A, it holds that*

$$\mathbb{P}(\Lambda(e) \mid A) = \frac{1}{c_{\mathcal{T}}(\theta, \lambda|_{R_A})} \frac{\mathrm{d}c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\lambda_e}, \quad \mathbb{P}(\Lambda(e), A) = \theta_e \frac{\mathrm{d}c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\theta_e}. \quad (5.6.10)$$

Proof. First consider the equation on the left. Recall that $\mathbb{P}(\Lambda(e) \mid A) = \frac{\mathbb{P}(\Lambda(e),A)}{\mathbb{P}(A)}$. From Lemma 5.6.17 it follows that the denominator is as in (5.6.10). For the numerator note that

$$\frac{\mathrm{d}c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\lambda_e} = \frac{\mathrm{d}\sum_{r \in R_A} \prod_{e \in E(r)} \lambda_e(r)\theta_e}{\mathrm{d}\lambda_e} = \sum_{r \in \Lambda(e) \cap R_A} \prod_{a \in E(r)} \lambda_a\theta_a = \mathbb{P}(\Lambda(e), A),$$

from which the equation on the left follows. Finally note that

$$\theta_e \frac{\mathrm{d}c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\theta_e} = \frac{\mathrm{d}c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\lambda_e},$$

since λ_e is equal to either one or zero, depending on whether a root-to-leaf path in R_A goes through e or not. $\qquad\square$

Notice that the derivatives of tree polynomials have the exact same interpretation of the ones of BNs as in Darwiche [2003]. Here we restricted our attention to square free staged trees but analogous results hold in the generic case: each monomial with indeterminates λ_e and θ_e of degree higher than one would need to be differentiated a number of times equal to the degree of that indeterminate.

Proposition 5.6.20. *In the notation of Proposition 5.6.19, the following holds for $e, e_1, e_2 \in E$:*

$$\mathbb{P}(\Lambda(e_1), \Lambda(e_2) \mid A) = \frac{1}{c_{\mathcal{T}}(\theta, \lambda|_{R_A})} \frac{\partial^2 c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\partial\lambda_{e_1}\partial\lambda_{e_2}}, \quad (5.6.11)$$

$$\mathbb{P}(\Lambda(e_1), \Lambda(e_2), A) = \theta_{e_1}\theta_{e_2} \frac{\partial c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\partial\theta_{e_1}\partial\theta_{e_2}}, \quad (5.6.12)$$

$$\mathbb{P}(A \mid \Lambda(e)) = \frac{\partial^2 c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\partial\theta_e\partial\lambda_e}. \quad (5.6.13)$$

Proof. Equations (5.6.11) and (5.6.12) follow applying twice equation (5.6.10). To prove (5.6.13), note that

$$\mathbb{P}(A \mid \Lambda(e)) = \frac{\mathbb{P}(A, \Lambda(e))}{\mathbb{P}(\Lambda(e))} = \frac{\mathrm{d}c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\lambda_e} \frac{1}{\theta_e} = \theta_e \frac{\mathrm{d}c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\theta_e} \frac{1}{\theta_e} = \frac{\mathrm{d}^2 c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\theta_e \mathrm{d}\lambda_e}.$$

$\square$

It is easy to then deduce from Proposition 5.6.20 the probabilistic meaning of higher order derivatives.

Importantly, in the staged tree model class derivatives can be associated to causal propositions. Note that such a result does not hold in general for the polynomials describing BN probabilities.

Proposition 5.6.21. *Suppose the staged tree is believed to be causal as in Thwaites et al. [2010]. Then in the notation of Proposition 5.6.20,*

$$\mathbb{P}(A \parallel \Lambda(e)) = \frac{\partial^2 c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\partial \theta_e \partial \lambda_e}, \tag{5.6.14}$$

where $\mathbb{P}(A \parallel \Lambda(e))$ is the probability of the event A when the system is forced to go through the edge e.

Proof. This follows noting that in a causal framework $\mathbb{P}(A \parallel \Lambda(e)) = \mathbb{P}(A \mid \Lambda(e))$. $\square$

Note that all the quantities in (5.6.10)-(5.6.14) can be used in sensitivity analysis. For instance we could investigate the changes in probability estimates when the system is set to be in a certain scenario of interest.

Example 5.6.22. We now compute a set of derivatives of the interpolating polynomial $c_{\mathcal{T}}$ in (5.6.7) with respect to λ_{31} and θ_{31} to perform probabilistic inference over the event A of political disruption. Thus, we consider the edge $e = (v_3, v_7)$ and

$$\frac{1}{c_{\mathcal{T}}(\theta, \lambda|_{R_A})} \frac{\mathrm{d}c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\lambda_e} = \frac{\theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{12}\theta_{31}}{\theta_{01}\theta_{11}\theta_{31} + \theta_{01}\theta_{12}\theta_{31}} = 1, \tag{5.6.15}$$

$$\theta_e \frac{\mathrm{d}c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\theta_e} = \theta_{13}(\theta_{01}\theta_{11} + \theta_{01}\theta_{12}) = \mathbb{P}(A), \tag{5.6.16}$$

$$\frac{\mathrm{d}^2 c_{\mathcal{T}}(\theta, \lambda|_{R_A})}{\mathrm{d}\theta_e \mathrm{d}\lambda_e} = \theta_{01}\theta_{11} + \theta_{01}\theta_{12}. \tag{5.6.17}$$

Observe that equation (5.6.15) is equal to one since every path associated to the event A must go through e. From the same argument it follows that equation (5.6.16) is equal to $\mathbb{P}(A)$. Equation (5.6.17) is a simple consequence of Bayes theorem, which can be checked algebraically.
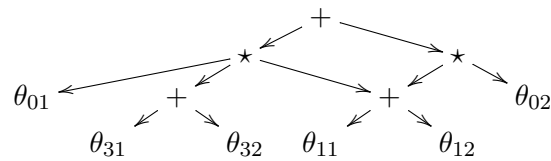
FIGURE 5.2: The arithmetic circuit of the polynomial in equation (5.6.7).

5.6.2.3 Trees as Circuits. The previous sections have introduced a comprehensive symbolic inferential toolbox for trees, based on the computation of the interpolating polynomial and its derivatives. In Darwiche [2003] it is shown that an efficient method to compute such polynomials is by representing them as an **AC**. This is a DAG whose leaves are the indeterminates and the inner nodes are labelled by multiplication and summation operations. The *size* of the circuit is measured by its number of edges.

ACs of staged tree polynomials are smaller in size than the ones associated to BNs for two reasons: first, a tree might have fewer root-to-leaf paths (as in our example); second, there can be less indeterminates because unfoldings with probability zero are not included in the model and coloured labels further decrease the number of indeterminates. Therefore, in asymmetric settings we can expect computations to be much faster for trees than BNs.

A major problem in the compilation of BNs polynomials consists of the identification of the AC of smallest size. This usually entails the computation of the BN's jointree and the application of rather complicated algorithms [Darwiche, 2003]. We note here that in tree models this is straightforward since the interpolating polynomial is naturally factored.

Example 5.6.23. Notice that equation (5.6.8) can be rewritten as $c_{\mathcal{T}}(\theta) = \theta_{01}(\theta_{11} + \theta_{12})(\theta_{31} + \theta_{31}) + \theta_{02}(\theta_{11} + \theta_{12})$. This gives us the AC in Figure 5.2 where leaves with the same parent are labelled by primitive probabilities from the same floret, and labels belonging to leaves in the tree are first summed in the AC. It is easy to deduce that the AC associated to the BN's polynomial in equation (5.6.9) would be much larger than the one in Figure 5.2. We note also that, whilst all the ACs deriving from BNs in Darwiche [2003] are trees, ours are more generally DAGs. This is a consequence of the more flexible stage structure associated to generic staged trees than the one of trees depicting BNs.

5.7 Conclusions

In this chapter, after discussing the links between algebraic and symbolic inferential methods, we have developed formal algebraic methods to investigate adequacy in IDSSs.

We have defined new separation conditions consisting of certain polynomial relations and demonstrated that these can guarantee adequacy and distributivity in general and in a variety of examples. In particular, the application of such methods to the class of BN models defined as DAG linear regressions lead us to deduce new formulae for the computation of moments of not only additive, but also multilinear, functions of random vectors over an agreed DAG. We have then extended current symbolic inferential methods to IDs and staged trees. For IDs we have defined a new symbolic representation of utilities, introduced a new symbolic evaluation algorithm and discussed asymmetric constraints in symbolic frameworks. Concerning staged trees, we have developed an inferential approach based on the computation of the derivatives of probability polynomials.

Chapter 6

Conclusions

In this last chapter we first briefly summarise the main contributions of this work. A more detailed discussion of the relevance of the results of the thesis is reported within each of the previous chapters. Importantly, this thesis gives a number of possible research directions that we discuss below. There are still many research avenues to explore and we outline here some issues we are currently addressing.

Summary of the Contributions of the Thesis

In this thesis we have defined the statistical framework of the IDSS that coherently aggregates the beliefs and preferences of separate panels of experts having different expertise. We discussed in Chapter 3 when and how it is possible to coherently combine both expert judgements and models into a unique entity and provided a variety of examples of IDSSs. In Chapter 4 we then defined new propagation algorithms for the computation of the expected utility of the policies supported by an IDSS. We developed in Chapter 5 new separation conditions for the panels' beliefs based on the algebraic form of the IDSS expected utility. In that chapter we further discussed symbolic approaches for decision making and inference in IDs and staged trees respectively.

In more details:

- Theorems 3.4.3 and 3.4.5, Lemma 3.4.7 and the results in Section 3.6 show that the framework of IDSSs allows for distributed Bayesian inferences, generalizing the standard Bayesian paradigm reviewed in Section 2.1. Specifically, we derived in Section 3.6 independence conditions for a variety of graphical models which generalize the standard independences over the parameter set reviewed in Chapter 2;

- the expert judgement combinations of Faria and Smith [1997] assume each expert has responsibility over the whole vertex set of an agreed BN or PCG model (see Proposition 2.6.5). In Proposition 3.6.5 we generalized their combination rules to cases where the experts partition the vertex set of an agreed graphical model;

- Lemma 3.5.2 and Theorem 3.5.5, using the newly notion of the natural comparator $\boldsymbol{d}^0$, generalized the standard Perlean notion of causality reviewed in Section 2.3.4 and introduced new inferential routines. As specified in the above mentioned results, the classic CBN model can be seen of an instance of an IDSS equipped with the natural comparator;

- in Definition 4.1.8 we introduced the new class of compatible utility factorization which exploits both generalized additive and utility independence, and can be represented by a specific subclass of the utility diagrams reviewed in Section 2.4.3. This new class of utility models is then showed to enable for the distributed computation of expected utilities in Theorem 4.1.13, which is also the first result to our knowledge using the idea of partial utility independence in expected utility computations;

- Section 4.3 then presents a variety of algorithms to compute these expected utilities in different situations, e.g. dynamic and non-dynamic. We further provided a long list of examples of these computations in practice in Section 4.4;

- we then focused in Chapter 5 on the computation of expected utilities for particular types of Bayesian network models and introduced a new procedure to quickly compute these scores in closed form. As shown in Lemma 5.5.6 and Proposition 5.5.10 the new concept of algebraic substitution together with the application of the multinomial theorem allowed us to instantly derive a polynomial form for expected utilities associated to BN models;

- the independence conditions for the distributed multi-expert inference routines in IDSSs were shown to be too strong for the computations of the expected utilities of the point above; we therefore identified in Theorems 5.5.8 and 5.5.13 minimal sets of independences for the computations of these expected utilities;

- lastly, we introduced two new symbolic inferential frameworks: one for IDs, where we symbolically defined utility parameters and deduced a polynomial expression for expected utilities; and one for CEGs and staged trees, where we showed that by computing certain derivatives of the polynomial defining the model, a user can answer a variety of inferential queries.

Future Work

Missing Data and Approximated Systems

We showed in Chapter 3 that evidence coming from certain observational and experimental studies can be introduced in the IDSS whilst retaining the distributivity of the systems. However often it could only be available data whose likelihood is not panel separable. This might be for example due to missing observations. We note here that there are two related practical solutions to this problem:

- panels can accommodate only a subset of the data into the system respecting panel separability. This gives the basis for a framework to analyse the system;

- methods can be developed where the distributivity property can only be approximatively satisfied.

In the latter case we would then need to introduce measures to quantify the effects of both the approximation of distributivity and the loss of data. One such measure could be for example the variation in expected utility defined as

$$|\bar{u}(\boldsymbol{d} \mid \boldsymbol{x}) - \bar{u}(\boldsymbol{d} \mid \boldsymbol{x}')|,$$

where $\boldsymbol{x}$ is a dataset with non panel separable likelihood, $\boldsymbol{x}'$ is a subset of $\boldsymbol{x}$ for which the panel separability property holds and $\bar{u}(\boldsymbol{d} \mid \boldsymbol{x})$ is the expected utility after the introduction of data. If the above measure is small, then the extent of not including observations that break distributivity will be almost irrelevant. This measure could then be embedded into a formal methodology of approximated decision support.

Non-Bayesian Approaches

The Bayesian framework provides a coherent and established methodology to model and reason about uncertainty. Therefore we claim that the collective should always strive to define their beliefs as subjective probability statements. However there are cases where members of panels will simply not agree on a unique probability specification. We could envisage the use of a pooling operator to aggregate these beliefs, but again the experts might disagree on the weights of such operator. There is therefore the need to develop inferential methods within IDSSs outside of the Bayesian formalism to deal with these situations.

Belief functions [Dempster, 1968, Shafer, 1976] and imprecise probabilities [Walley, 1991] represent possible candidates, since these do not require the experts to agree on a unique specification. In particular *credal networks* [Cozman, 2000] can be used as an integrating structure in these frameworks. These are DAGs defining for each vertex a credal set, a set of probability distributions, instead of one only distribution as in BNs.

Real World Application of IDSSs

The focus of this thesis has been on developing a formal theory to aggregate expert judgements in complex settings, as nuclear emergency management. Since a formal theory is now in place we can start looking at developing these types of systems to support actual decision making. One obvious domain of application would be nuclear emergency management since, as we noted, Bayesian models for many of the modules of any plausible network have been already defined. We are currently starting to develop an IDSS to deal with food security issues. In this direction, Barons et al. [2014] built a DBN to model the production of sugar.

Propagation in Group CEGs

The propagation algorithms we defined in Chapter 4, although working for a wide variety of commonly used models, do not allow for asymmetric conditional independence statements to be explicitly represented. It would be therefore relevant to develop algorithms for models exhibiting such an asymmetric structure, as for example CEGs. In Section 3.6.1.7 we showed that these models can be part of the structural consensus of an IDSS. We can therefore now build on the algorithm of Thwaites et al. [2008] to take into account the underlying multi-expert structure of the graph.

Dynamic and Multivariate Algebraic Substitution

Algebraic substitutions have proved extremely powerful in both computing conditional expected utilities and moments of functions over an underlying DAG. These substitutions work for non-dynamic models whose vertices are univariate random variables. Given the often geographic nature of the problems IDSSs aim at addressing, there is a need to generalise such recursions to the multivariate case. A natural framework where this could be achieved is by representing probabilities as *tensors* [see e.g. McCullagh, 1987, for a discussion of their use in statistics]. Furthermore algebraic substitutions can provide the basis for moment forecasting in 'polynomial MDM models', generalising the recursions of Queen et al. [2008] that forecast the covariances of an LMDM.

Symbolic Continuous IDs

In Section 5.6.1 we introduced a symbolic definition of IDs for discrete decisions and variables. This formalism can be rather straightforwardly extended to continuous IDs. A symbolic evaluation algorithm for these models would then use the idea of algebraic substitution to compute expected utilities of decision nodes. Madsen and Jensen [2005] developed, in a non-symbolic framework, an evaluation algorithm for continuous IDs embedding additive utility factorisations with quadratic marginal utilities functions. This can easily identify optimal decisions since the associated expected utility has a unique maximum. We have started developing similar recursions in a very wide class of IDs with multilinear and polynomial utilities, where in addition a compliance error is associated to every decision node. Just as in the discrete case, we can then implement such an algorithm in a computer algebra system.

Asymmetric IDs

In Section 5.6.1 we briefly discussed the imposition of asymmetric constraints in IDs. We also noted that great computational advantages would be achieved if these constraint were implemented during the evaluation process. We are planning to develop such a methodology by starting looking at *decision circuits* [Bhattacharjya and Shachter, 2007, 2012], a graphical representation of asymmetric problems through a variation of an AC.

Computer Algebra and Sensitivity Analysis in Trees

The symbolic definition of probabilities of BN models has given rise to a large body of literature on formal symbolic sensitivity analyses [see e.g. Chan and Darwiche, 2001, 2004]. Having symbolically defined probabilities in staged trees, we can now start developing sensitivity analyses techniques tailored to the structure of the staged tree polynomials. Alongside these theoretical advances, we are looking to implement the symbolic definition of staged trees in a computer algebra system to widen the use of such models in practice.

Appendix A

Proofs

A.1 Proofs of Chapter 2

A.1.1 Proof of Proposition 2.5.10

We develop the proof via backward induction over the random and decision vertices of the MID, starting from Y_n. Define, for $i \in [n]$,

$$\bar{u}^i = \int_{\mathcal{Y}_{[n]^{\mathbb{V}}_{i-1}}} \max_{\mathcal{Y}_{[n]^{\mathbb{D}}_{i-1}}} \sum_{I \in \mathcal{P}_0([m])} h^{n_I - 1} \prod_{i \in I} k_i u_i(\boldsymbol{y}_{P_i}) f\left(\boldsymbol{y}_{[n]^{\mathbb{V}}_{i-1}} \mid \boldsymbol{y}_{\Pi_{[n]^{\mathbb{V}}_{i-1}}}\right) \mathrm{d}\boldsymbol{y}_{[n]^{\mathbb{V}}_{i-1}},$$

where $[n]^{\mathbb{V}}_{i-1} = [n] \setminus [i-1] \cap \mathbb{V}$, $[n]^{\mathbb{D}}_{i-1} = [n] \setminus [i-1] \cap \mathbb{D}$ and $\Pi_{[n]^{\mathbb{V}}_{i-1}} = \cup_{j \in [n]^{\mathbb{V}}_{i-1}} \Pi_j$.

The DM's preferences are a function of Y_n only through $k_m u_m(\boldsymbol{y}_{P_m})$, since by construction $n = j_m \in \mathbb{J}$. Therefore, this quantity can be either maximised or marginalised as in equation (2.5.2) to compute $\bar{u}_n(\boldsymbol{y}_{B_n})$. Note that B_n includes only the indices of the variables $\bar{u}_n$ formally depends on, since $B_n = P_m \setminus n$, if $n \in \mathbb{D}$, whilst $B_n = P_m \cup \Pi_n \setminus n$, if $n \in \mathbb{V}$. Then

$$\bar{u}^n = \sum_{I \in \mathcal{P}_0([m])} h^{n_I - 1} \prod_{i \in I} (\mathbb{1}_{\{i \neq n\}} k_i u_i(\boldsymbol{y}_{P_i}) + \mathbb{1}_{\{i = n\}} \bar{u}_i(\boldsymbol{y}_{B_i})).$$

Now consider Y_{n-1}. If $n - 1 \notin \mathbb{J}$, then $\bar{u}^n$ is a function of Y_{n-1} only through $\bar{u}_n$. Therefore maximisation and marginalisation steps can be computed as in equation (2.5.5) to compute $\bar{u}_{n-1}(\boldsymbol{y}_{B_{n-1}})$. Again B_{n-1} includes the indices of the variables $\bar{u}_{n-1}$ formally depends on, since $B_{n-1} = P_m \setminus \{n, n-1\}$, if $n, n-1 \in \mathbb{D}$, $B_{n-1} = P_m \cup \Pi_n \cup \Pi_{n-1} \setminus \{n, n-1\}$, if $n, n-1 \in \mathbb{V}$, $B_{n-1} = P_m \cup \Pi_{n-1} \setminus \{n, n-1\}$, if $n \in \mathbb{D}$ and $n-1 \in \mathbb{V}$, $B_{n-1} =$

$P_m \cup \Pi_n \setminus \{n, n-1\}$, if $n \in \mathbb{V}$ and $n-1 \in \mathbb{D}$. Then

$$\bar{u}^{n-1} = \sum_{I \in \mathcal{P}_0([m])} h^{n_I - 1} \prod_{i \in I} (\mathbb{1}_{\{i \neq n\}} k_i u_i(\boldsymbol{y}_{P_i}) + \mathbb{1}_{\{i=n\}} \bar{u}_{i-1}(\boldsymbol{y}_{B_{i-1}})).$$

Conversely, if $n - 1 \in \mathbb{J}$, $\bar{u}^n$ is potentially a function of Y_{n-1} through both $u_{m-1}(\boldsymbol{y}_{P_{m-1}})$ and $\bar{u}_n(\boldsymbol{y}_{B_n})$ and note that $\bar{u}^n$ can be written in this case as

$$\sum_{I \in \mathcal{P}_0([m-2])} h^{n_I - 1} \prod_{i \in I} k_i u_i + u'_{m-1} + \left(\sum_{i \in \mathcal{P}_0([m-2])} h^{n_i - 1} \prod_{i \in I} k_i u_i \right) u'_{m-1},$$

where

$$u'_{m-1} = h k_{m-1} u_{m-1}(\boldsymbol{y}_{P_{m-1}}) \bar{u}_n(\boldsymbol{y}_{B_n}) + k_{m-1} u_{m-1}(\boldsymbol{y}_{P_{m-1}}) + \bar{u}_n(\boldsymbol{y}_{B_n}).$$

Therefore, optimisation and marginalisation steps can be performed over u'_{m-1} as specified in equations (2.5.3) and (2.5.4) respectively. Then note that $\bar{u}^{n-1}$ can be written as

$$\begin{aligned}
\bar{u}^{n-1} &= \sum_{I \in \mathcal{P}_0([m-2])} h^{n_I - 1} \prod_{i \in I} k_i u_i + \bar{u}_{n-1} + \left(\sum_{i \in \mathcal{P}_0([m-2])} h^{n_i - 1} \prod_{i \in I} k_i u_i \right) \bar{u}_{n-1} \\
&= \sum_{I \in \mathcal{P}_0([m-1])} h^{n_I - 1} \prod_{i \in I} (\mathbb{1}_{\{i \neq n-1\}} k_i u_i(\boldsymbol{y}_{P_i}) + \mathbb{1}_{\{i=n-1\}} \bar{u}_i(\boldsymbol{y}_{B_i})).
\end{aligned}$$

Now for a $j \in [n-2]$ and assuming with no loss of generality that k is the index of a utility vertex such that $j_{k-1} < j \leq j_k$, we have that

$$\bar{u}^j = \sum_{I \in \mathcal{P}_0([k])} h^{n_I - 1} \prod_{i \in I} (\mathbb{1}_{\{i \neq j\}} k_i u_i(\boldsymbol{y}_{P_i}) + \mathbb{1}_{\{i=j\}} \bar{u}_i(\boldsymbol{y}_{B_i})).$$

Therefore, at the following step, when considering Y_{j-1}, we can proceed as done with Y_{n-1} by maximisation and marginalisation in equations (2.5.3)-(2.5.5) to compute $\bar{u}^{j-1}$. Thus, at the conclusion of the procedure, $\bar{u}_1$ yields the expected utility of the optimal decision.

A.2   Proofs of Chapter 3

A.2.1   Proof of Theorem 3.4.3

Fix a $\boldsymbol{d} \in \mathcal{D}$ and for simplicity suppress this dependence and the time index t. We need show that under the conditions of the theorem at any time t and under any policy the

SB will continue to hold panel independent beliefs, i.e. that, for $i \in [m]$,

$$\boldsymbol{\theta}_i \perp\!\!\!\perp \boldsymbol{\theta}_{i-} \mid I_+, \tag{A.2.1}$$

and that when assessing $\boldsymbol{\theta}_i$, she will only use the information G_i would use if acting autonomously in assessing the information she needs to deliver, i.e.

$$\boldsymbol{\theta}_i \perp\!\!\!\perp I_+ \mid I_{CK}, I_{ii}. \tag{A.2.2}$$

This is then sufficient for soundness and distributivity. Because even if all panellists could share each other's information then, given all panel beliefs, they would come to the same assessment about the joint distribution of the relevant parameters: that all panel subvectors are mutually independent of each other and that these margins will simply be the margins of the associated panel were they making decisions autonomously.

The proof simply uses the semi-graphoid axioms of conditional independence in Proposition 2.3.2. We start by proving the panel independence condition in (A.2.1). Note that from common separability in equation (3.4.4) it follows that

$$\boldsymbol{\theta}_i \perp\!\!\!\perp \boldsymbol{\theta}_{i-} \mid I_{CK}.$$

which combined with the separately informed condition in equation (3.4.2) through strong decomposition and using the symmetric property of semi-graphoids axioms gives

$$\boldsymbol{\theta}_{i-} \perp\!\!\!\perp I_{ii}, \boldsymbol{\theta}_i \mid I_{CK}.$$

Using again strong decomposition and symmetric arguments, it follows that

$$\boldsymbol{\theta}_i \perp\!\!\!\perp \boldsymbol{\theta}_{i-} \mid I_{ii}, I_{CK}. \tag{A.2.3}$$

Now the cutting condition in equation (3.4.3) together with equation (A.2.3) implies by strong decomposition that

$$\boldsymbol{\theta}_i \perp\!\!\!\perp I_*, \boldsymbol{\theta}_{i-} \mid I_{CK}, I_{ii}. \tag{A.2.4}$$

Then again by strong decomposition we have that

$$\boldsymbol{\theta}_i \perp\!\!\!\perp \boldsymbol{\theta}_{i-} \mid I_{CK}, I_{ii}, I_*.$$

Since I_{ii} is a function of I_* the above expression is equivalent to

$$\boldsymbol{\theta}_i \perp\!\!\!\perp \boldsymbol{\theta}_{i-} \mid I_{CK}, I_*. \tag{A.2.5}$$

Now the delegatable condition in equation (3.4.1) can be written as

$$I_+ \perp\!\!\!\perp \boldsymbol{\theta}_i, \boldsymbol{\theta}_{i-} \mid I_{CK}, I_*,$$

and using strong decomposition it follows that

$$I_+ \perp\!\!\!\perp \boldsymbol{\theta}_i \mid I_{CK}, I_*, \boldsymbol{\theta}_{i-}. \tag{A.2.6}$$

Combining via strong decomposition equations (A.2.5) and (A.2.6) and using the symmetry property we have that

$$\boldsymbol{\theta}_i \perp\!\!\!\perp I_+, \boldsymbol{\theta}_{i-} \mid I_{CK}, I_*.$$

Using again strong decomposition it follows that

$$\boldsymbol{\theta}_i \perp\!\!\!\perp \boldsymbol{\theta}_{i-} \mid I_{CK}, I_*, I_+,$$

which, since I_* and I_{CK} are functions of I_+, can be written as equation (A.2.1). This shows that panel independence directly follows from the four conditions of Definition 3.4.2.

To show that equation (A.2.2) holds, note that another implication of delegatability in equation (3.4.1) by strong decomposition is that

$$\boldsymbol{\theta}_i \perp\!\!\!\perp I_+ \mid I_{CK}, I_{ii}, I_*, \tag{A.2.7}$$

where again we used the fact that I_{ii} is a function of I_*. Now noting that by strong decomposition equation (A.2.4) implies that

$$\boldsymbol{\theta}_i \perp\!\!\!\perp I_* \mid I_{CK}, I_{ii}, \tag{A.2.8}$$

it follows from equations (A.2.7) and (A.2.8) by strong decomposition that

$$\boldsymbol{\theta}_i \perp\!\!\!\perp I_+, I_* \mid I_{CK}, I_{ii},$$

which, since I_* is a function of I_+ is equivalent to equation (A.2.2).

A.2.2 Proof of Theorem 3.4.5

Fix a policy $\boldsymbol{d} \in \mathcal{D}$ and suppress for ease of notation this dependence. Under the initial hypotheses, by Theorem 3.4.3

$$\perp\!\!\!\perp_{i \in [m]} \boldsymbol{\theta}_i \mid I_{CK}^0, I_*^0,$$

implying that the prior joint density can be written in a product form

$$\pi(\boldsymbol{\theta}) = \prod_{i \in [m]} \pi_i(\boldsymbol{\theta}_i).$$

It follows that under this admissibility protocol

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}^t) = \prod_{i \in [m]} \pi_i(\boldsymbol{\theta}_i, \boldsymbol{t}_i(\boldsymbol{x}^t)),$$

where from the form of the likelihood above

$$\pi_i(\boldsymbol{\theta}_i, \boldsymbol{t}_i(\boldsymbol{x}^t)) \propto l_i(\boldsymbol{\theta}_i \mid \boldsymbol{t}_i(\boldsymbol{x}^t))\pi_i(\boldsymbol{\theta}_i).$$

By hypothesis $\pi_i(\boldsymbol{\theta}_i, \boldsymbol{t}_i(\boldsymbol{x}^t))$ will be delivered by G_i and adopted by the SB. So the IDSS is sound. In particular we can deduce, through the definition of panel separability and the results of Theorem 3.4.3, that

$$\perp\!\!\!\perp_{i\in[m]}\boldsymbol{\theta}_i \mid I_{CK}^t, I_+^t \iff \perp\!\!\!\perp_{i\in[m]}\boldsymbol{\theta}_i \mid I_{CK}^t, I_*^t.$$

So we have separability a posteriori. Note that in the notation above

$$I_+^t = \left\{ I_{CK}^t, I_*^t \right\} = \left\{ I_{CK}^0, I_*^0, \boldsymbol{x}^t \right\},$$

and for $i \in [m]$

$$\left\{ I_{CK}^t, I_{ii}^t \right\} = \left\{ I_{CK}^t, I_{ii}^0, \boldsymbol{t}_i(\boldsymbol{x}^t) \right\}.$$

Since by definition $\boldsymbol{t}_i(\boldsymbol{x}^t)$ is known to G_i, $i \in [m]$, the system is also delegatable. Finally note that if

$$l(\boldsymbol{\theta} \mid \boldsymbol{x}^t) \neq \prod_{i\in[m]} l_i(\boldsymbol{\theta}_i \mid \boldsymbol{t}_i(\boldsymbol{x}^t))$$

on a set A of non zero prior measure, then the conditional density $\pi_A(\boldsymbol{\theta}|\boldsymbol{x}^t)$ on A will have the property that for all $\boldsymbol{\theta} \in A$

$$\pi_A(\boldsymbol{\theta}|\boldsymbol{x}^t) \neq \prod_{i\in[m]} \pi_{A,i}(\boldsymbol{\theta}_i, \boldsymbol{t}_i(\boldsymbol{x}^t)),$$

where $\pi_{A,i}$ denotes the density delivered by panel G_i for the parameters it oversees in A. This means that panel parameters are a posteriori dependent and so in particular the density determined by the margins is not sound.

A.2.3   Proof of Theorem 3.6.9

Suppress for simplicity the dependence on $\boldsymbol{d}$ and the time index t. Since the IDSS is a priori sound, the prior density can be written as

$$\pi(\boldsymbol{\theta}) = \frac{\prod_{C_i\in\mathcal{C}} \pi(\boldsymbol{\theta}_{C_i})}{\prod_{S\in\mathcal{S}} \pi(\boldsymbol{\theta}_S)},$$

where, by soundness, the distributions over the separators can be computed from the any distribution over a clique containing that separator. Now soundness a posteriori is

ensured iff the posterior can be uniquely written as

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{\prod_{C_i \in \mathcal{C}} \pi(\boldsymbol{\theta}_{C_i} \mid \boldsymbol{x})}{\prod_{S \in \mathcal{S}} \pi(\boldsymbol{\theta}_S \mid \boldsymbol{x})}, \tag{A.2.9}$$

where $\boldsymbol{x} = (\boldsymbol{x}_i^{\mathrm{T}})_{i \in [n]}^{\mathrm{T}}$. Since each panel updates its distribution using its own independent dataset, then it follows that the posterior over the cliques enjoys the factorisation at the numerator of (A.2.9). Calling $\pi_{i,S}^*$ the posterior distribution of panel G_i over S, we then have that

$$\pi_{i,S}(\boldsymbol{\theta}_S \mid \boldsymbol{x}) \propto l_i(\boldsymbol{\theta}_S \mid t_i(\boldsymbol{x}_{i,S}^t))\pi_i(\boldsymbol{\theta}_S).$$

Therefore, from the above equation, the posterior distribution can be written as in (A.2.9) iff, for every i such that $S \subset C_i$, the sufficient statistic over the separator is equal for every panel G_i.

A.3   Proofs of Chapter 4

A.3.1   Proof of Theorem 4.1.13

We develop this proof via backward induction both through the vertices of the DAG and through time. For the purpose of this proof define for $t = T$

$$\bar{u}^{T,i}(\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_{i-1}^T, \boldsymbol{y}_i^{T-1}, \ldots, \boldsymbol{y}_n^{T-1}, \boldsymbol{d}^T) = \int_{\mathcal{Y}_i} \cdots \int_{\mathcal{Y}_n} u^{\mathcal{G}} f_{T,i} \cdots f_{T,n} \mathrm{d}\boldsymbol{y}_i(T) \cdots \mathrm{d}\boldsymbol{y}_n(T), \tag{A.3.1}$$

and note that $\bar{u}^{T,1} \equiv \bar{u}^T$.

First, without any loss of generality, fix a policy $\boldsymbol{d}^T$. Then start the backward induction from $\boldsymbol{Y}_n(T)$, which, by construction, is a leaf of the time slice DAG at time T. For a leaf, $\boldsymbol{Y}_i(T)$ say, it follows from (4.1.5) that $\tilde{u}_{T,i} = u_i^{\mathcal{G}}(\boldsymbol{r}_{A_i})$ and note that consequently $u_i^{\mathcal{G}}$ is a function of $\boldsymbol{Y}_n(T)$ only through $\tilde{u}_{T,n}$. Therefore $\tilde{u}_{T,n}$ can then be simply marginalised as in equation (4.1.6) to obtain $\bar{u}_{T,n}$. Furthermore

$$\bar{u}^{T,n} = \sum_{i \in \{Le \setminus \{n\}\}} u_i^{\mathcal{G}}(\boldsymbol{r}_{A_i}) + \bar{u}_{T,n}\left(\boldsymbol{y}_{A_n'}^T, \boldsymbol{y}_n^{T-1}, \boldsymbol{d}^T\right). \tag{A.3.2}$$

Now consider $\boldsymbol{Y}_{n-1}(T)$. The vertex associated with this random vector in the time slice DAG is either the father of $\boldsymbol{Y}_n(T)$ or a leaf of the DAG. In the latter case, since by construction $n - 1 \in U$, the exact same method followed for $\boldsymbol{Y}_n(T)$ can be applied to

$\boldsymbol{Y}_{n-1}(T)$, and thus

$$\bar{u}^{T,n-1} = \sum_{i \in \{Le \setminus \{n,n-1\}\}} u_i^{\mathcal{G}}(\boldsymbol{r}_{A_i}) + \sum_{j=n-1}^{n} \bar{u}_{T,j} \left(\boldsymbol{y}_{A_j'}^T, \boldsymbol{y}_j^{T-1}, \boldsymbol{d}^T \right). \tag{A.3.3}$$

If on the other hand $\boldsymbol{Y}_{n-1}(T)$ is the father of $\boldsymbol{Y}_n(T)$, then by construction $\boldsymbol{Y}_{n-1}(T)$ has only one son. Thus from equation (4.1.5) $\bar{u}_{T,n} \equiv \tilde{u}_{T,n-1}$ and equation (A.3.2) is a function of $\boldsymbol{Y}_{n-1}(T)$ only through $\bar{u}_{T,n}$. In order to deduce $\bar{u}^{T,n-1}$ only $\tilde{u}_{T,n-1}$ has to be marginalised with respect to $f_{T,n-1}$ and therefore

$$\bar{u}^{T,n-1} = \sum_{i \in \{Le \setminus \{n,n-1\}\}} u_i^{\mathcal{G}}(\boldsymbol{r}_{A_i}) + \bar{u}_{T,n-1} \left(\boldsymbol{y}_{A_{n-1}'}^T, \boldsymbol{y}_{n-1}^{T-1}, \boldsymbol{y}_n^{T-1}, \boldsymbol{d}^T \right). \tag{A.3.4}$$

We can note from equations (A.3.3) and (A.3.4) that $\bar{u}^{T,n-1}$ consists of the linear combination of two summations: the first over the leaves of the graphs with index j smaller than $n-1$ of utility terms $u_j^{\mathcal{G}}$; the second over the indices j bigger or equal than $n-1$ of the terms $\bar{u}_{T,j}$ such that the father of $\boldsymbol{Y}_j(T)$ has an index smaller than $n-1$ in the time slice DAG. So for example in equation (A.3.3) the second summation is over both n and $n-1$ since the associated vertices are both leaves of the graphs. On the other hand in equation (A.3.4) there is no term $\bar{u}_{T,n}$ since its father has index $n-1$. More generally, for $j \in [n]$, $\bar{u}^{T,j}$ can be written as the linear combination of the following two summations:

- the first over the indices i in $Le \cap [j-1]$ of $u_i^{\mathcal{G}}$;

- the second over the indices k in $B_j = \{k \geq j : F_k < j\}$ of $\bar{u}_{T,k}$, where F_k is the index of the father of $\boldsymbol{Y}_k^{\mathrm{T}}$.

Therefore, for a $j \in [n]$, we have that

$$\bar{u}^{T,j} = \sum_{i \in \{Le \cap [j-1]\}} u_i^{\mathcal{G}}(\boldsymbol{r}_{A_i}) + \sum_{k \in B_j} \bar{u}_{T,k} \left(\boldsymbol{y}_{A_k'}^T, \boldsymbol{y}_k^{T-1}, \boldsymbol{y}_{Dn_k}^{T-1}, \boldsymbol{d}^T \right), \tag{A.3.5}$$

where Dn_k is the set of the indices of the descendants of $\boldsymbol{Y}_k^{\mathrm{T}}$.

In particular for $\boldsymbol{Y}_2(T)$ we can write equation (A.3.5) as

$$\bar{u}^{T,2} = \sum_{k \in S_1} \bar{u}_{T,k} \left(\boldsymbol{y}_1^T, \boldsymbol{y}_k^{T-1}, \boldsymbol{y}_{S_k}^{T-1}, \boldsymbol{d}^T \right), \tag{A.3.6}$$

since, by the connectedness of the time slice DAG, $\boldsymbol{Y}_1(T)$ is the father of all the vertices whose father's index is not $[n] \setminus \{1\}$. It then follows that equation (A.3.6) corresponds

to $\tilde{u}_{T,1}$, as defined in equation (4.1.5), and therefore $\bar{u}^T$ can be written as in equation (4.1.4). Thus Theorem 4.1.13 holds for time T.

Now, since $\boldsymbol{Y}_1(T)$ is the unique root of the time slice DAG, if $i, j \in S_1$, then

$$A_i' \cap A_j' = \{1\}. \tag{A.3.7}$$

Suppose that any vertex $\boldsymbol{Y}_j(T)$, for $j \in S_1$, is either connected by a path to one only leaf of the DAG or is a leaf of the graph itself. Because of the identity in equation (A.3.7) and because of the algebraic form of equation (A.3.6), which consists of a linear combination of the terms $\bar{u}_{T,j}$, for $j \in S_1$, we can deduce that equation (4.1.7) holds for the last time slice. Now, consider the case where one vertex $\boldsymbol{Y}_j(T)$ with index in S_1 is connected to more than one leaf. Equation (4.1.5) guarantees the existence of a vertex $\boldsymbol{Y}_i(T)$, $i > j$, connected to both $\boldsymbol{Y}_j(T)$ and the above mentioned leaves, such that $\tilde{u}_{T,i}$ can be written as a linear combination of terms $\bar{u}_{T,k}$, for which each of these terms is a function of one of the leaves only. It therefore follows that equation (4.1.7) also holds in this case.

Therefore equation (4.1.7) guarantees that $\bar{u}^{T,1}$ can be written as a linear combination of terms involving only variables in the same ancestral set. Since also the probability factorisation does not change as formalised in Proposition 4.1.2, the exact same recursions we explicated at time T can then be followed at time $T-1$ by substituting $u_i^{\mathcal{G}}$ with $\hat{u}_{T-1,i}$, $i \in Le$, in equations (A.3.1)-(A.3.5) and by changing the time index. This then also holds for any time slice t, $1 \leq t \leq T-1$, since $\bar{u}^{t,1}$ will be again a linear combination of terms $\hat{u}_{t-1,i}$, $i \in Le$, and the probability density function factorises as in Proposition 4.1.2.

A.3.2 Proof of Theorem 4.3.2

To prove Theorem 4.3.2 we proceed as follows:

- We relate the lines of the pseudo-code of Algorithm 4.3.1 to the equations (4.1.4)-(4.1.7) of Theorem 4.1.13 and their variations which include optimisation steps in equations (4.3.1) and (4.3.2);

- We then show that each panel and the SB have sufficient information to perform the steps of the algorithm they are responsible for;

- We conclude by showing that the optimisation steps, which in the algorithm correspond to lines (8) and (15), are able to identify optimal decisions using only combinations of quantities individual panels are able to calculate.

We start with the first two bullets. Line (1) describes the backward induction step over the time index, t, whilst line (2) does the same over the index of the vertices of the graph, i. Now note that in lines (5)-(7), Panel $G_i : \tilde{u}_{t,i}$ using equation (4.1.5). Each panel has enough information to do this, since line (10) guarantees that the scores are communicated to the panels overseeing father vertices and line (14) denotes the fact that the SB transmits $\hat{u}_{t,i}$ to the appropriate panels. The functions $\tilde{u}_{t,i}$ are then sent to the SB, who performs an optimisation step in line (8) and communicates the result back to the panel. We address the validity of this step below.

Since the $SB : u_{t,i}^* \longrightarrow G_i$, each panel is able to compute $\bar{u}_{t,i}^*$ (lines 10-11) following equation (4.3.2). As noted before, if i is not the root of the DAG, $\bar{u}_{t,i}^*$ is sent to the appropriate panel, whilst, if $i = 1$, as specified by the if statement in line (9), $\longrightarrow SB$. For each time slice with time index $t \neq 1$, lines (13)-(14) compute $\hat{u}_{t,i}$, as in equation (4.1.7). These are sent to the appropriate panels, which can then continue the backward inductive process from the time slice with a lower time index. If on the other hand $t = 1$, then the expected utility is a function of the initial decision space $\mathcal{D}(0)$ only. The SB can then perform a final optimisation step over this space and thus conclude the algorithm (line 15).

We now address the optimisation steps. The influence on the scores associated with time slices with index bigger than t of a decision space $\mathcal{D}_i(t)$ are included, by construction, only in the terms $\hat{u}_{t,k}$, where k is either the index of a descendant $\boldsymbol{Y}_k(t)$ of $\boldsymbol{Y}_i(t)$ or $k = i$. Further note that the same decision space $\mathcal{D}_i(t)$ can affect the scores of terms including descendants of $\boldsymbol{Y}_i(t)$ at the same time point. Thus the whole contribution of $\mathcal{D}_i(t)$ is summarised within $\tilde{u}_{t,i}$, as it can be seen by recursively using equations (4.1.5) and (4.1.6).

Now, as specified by equation (4.3.1), the optimisation step over $\mathcal{D}_i(t)$ is performed by maximising $\tilde{u}_{t,i}$, which carries all the information concerning this decision space. More specifically, no other term is an explicit function of $\mathcal{D}_i(t)$ at this stage of the algorithm, as guaranteed by equations (4.1.2). Finally, Structural Assumption 4.1.4 guarantees that all the elements that appears as arguments of $\tilde{u}_{t,i}$ are observed and therefore known at the time the decision associated to this decision space needs to be made.

A.4 Proofs of Chapter 5

A.4.1 Proof of Theorem 5.4.6

Fix a policy $\boldsymbol{d} \in \mathcal{D}$ and suppress this dependence. Since $\boldsymbol{R} = \boldsymbol{Y}$, the utility factorisation is panel separable and the marginal utilities are polynomials, the utility function can be written as

$$u(\boldsymbol{y}) = \sum_{I \in \mathcal{P}_0([m])} k_I \sum_{i \in I} \left(\sum_{b_i \in B_i} \rho_{b_i} y_i^{b_i} \right), \tag{A.4.1}$$

for $B_i \subset \mathbb{Z}_{>0}$. Note also that we can rewrite (A.4.1) as

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

where

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

$$\hat{u}(y_m) = \sum_{I \in \mathcal{P}_0([m]) \cap \{m\}} k_I \prod_{i \in I} \left(\sum_{b_i \in B_i} \rho_{b_i} y_i^{b_i} \right). \tag{A.4.2}$$

Calling $\boldsymbol{\theta}$ the overall parameter vector of the IDSS, the conditional expected utility function, $\mathbb{E}(u(\boldsymbol{Y}) \mid \boldsymbol{\theta})$, can be written applying sequentially the tower rule of expectation as

$$\mathbb{E}(u(\boldsymbol{Y}) \mid \boldsymbol{\theta}) = \mathbb{E}_{Y_1|\boldsymbol{\theta}} \left(\mathbb{E}_{Y_2|Y_1,\boldsymbol{\theta}} \left(\cdots \mathbb{E}_{Y_{m-1}|\boldsymbol{Y}_{[m-2]},\boldsymbol{\theta}} \left(\hat{u}(\boldsymbol{y}_{[m-1]}) + \mathbb{E}_{Y_m|\boldsymbol{Y}_{[m-1]},\boldsymbol{\theta}} \left(\hat{u}(y_m) \right) \right) \right) \right). \tag{A.4.3}$$

From equation (A.4.2), 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

$$E_{Y_m|\boldsymbol{Y}_{[m-1]},\boldsymbol{\theta}} \left(\hat{u}(\boldsymbol{y}_m) \right) = p_m(\boldsymbol{Y}_{[m-1]}, \boldsymbol{\theta}), \tag{A.4.4}$$

where p_m is a generic polynomial function. Thus $\hat{u}(\boldsymbol{Y}_{[m-1]}) + \mathbb{E}_{Y_m|\boldsymbol{Y}_{[m-1]},\boldsymbol{\theta}} \left(\hat{u}(\boldsymbol{y}_m) \right)$ is also a polynomial function of the same arguments.

Following the same reasoning to deduce (A.4.4), we then have that

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

where p_{m-1} is a generic polynomial function. Therefore the same procedure can be applied to all the expectations in (A.4.3). We can then write $\mathbb{E}(u(\boldsymbol{Y}) \mid \boldsymbol{\theta}) = p_1(\boldsymbol{\theta})$, where

p_1 is a generic polynomial function. This is by construction an algebraic conditional expected utility, where the functions λ_{ij} are monomials. Partial independence and Lemma 5.4.1 then guarantees score separability holds.

A.4.2 Proof of Proposition 5.5.9

We prove this result via induction over the number of vertices of the DAG. If the DAG has only one vertex

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

which can be seen as an instance of equation (5.5.11). Assume the result holds for a network with $n-1$ vertices. A multilinear utility factorisation can be rewritten as

$$u(\boldsymbol{r}) = \sum_{I \in \mathcal{P}_0([n-1])} k_I \prod_{i \in I} u_i(y_i) + \sum_{I \in \mathcal{P}_0([n]) \cap \{n\}} k_I \prod_{i \in I \setminus \{n\}} u_i(y_i) u_n(y_n) + k_n u_n(y_n). \quad (A.4.5)$$

The first term on the rhs of (A.4.5) is by inductive hypothesis equal to the sum of all the possible monomial of degree $\boldsymbol{\alpha}^{\mathrm{T}} = (\alpha_1, \ldots, \alpha_{n-1}, 0)$ where $0 < \alpha_i < n_i$, $i \in [n]$. The other terms only include monomials such that the exponent of y_n is not zero. We have that, letting $\boldsymbol{n}_{i-1}^{\mathrm{T}} = (n_i)_{i \in [n-1]}$ and $\boldsymbol{y}_{[n-1]} = \prod_{i \in [n-1]} y_i$,

$$\sum_{I \in \mathcal{P}_0([n]) \cap \{n\}} k_I \prod_{i \in I \setminus \{n\}} u_i(y_i) u_n(y_n) + k_n u_n(y_n) = \sum_{\boldsymbol{0} <_{lex} \boldsymbol{\alpha} \leq_{lex} \boldsymbol{n}_{n-1}} c_{\boldsymbol{\alpha}} \boldsymbol{y}_{[n-1]}^{\boldsymbol{\alpha}} \left(\sum_{i \in [n_n]} \rho_{ni} y_n^i \right) + k_n u_n(y_n)$$

$$= \sum_{\substack{\boldsymbol{0} <_{lex} \boldsymbol{\alpha} \leq_{lex} \boldsymbol{n}_{n-1} \\ i \in [n_n]}} c_{\boldsymbol{\alpha}} \rho_{ni} \boldsymbol{y}_{[n-1]}^{\boldsymbol{\alpha}} y_n^i + k_n u_n(y_n)$$

$$= \sum_{\substack{\boldsymbol{0}' <_{lex} \boldsymbol{\alpha} \leq_{lex} \boldsymbol{n}_n \\ \alpha_n \neq 0}} c_{\boldsymbol{\alpha}} \boldsymbol{y}_{[n]}^{\boldsymbol{\alpha}}. \quad (A.4.6)$$

Therefore, the sum of the first term on the rhs of (A.4.5) and (A.4.6) equals equation (5.5.11).

A.4.3 Proof of Theorem 5.6.3

For a subset $I \in \mathcal{P}_0([m])$, let j_I be the index of the variable appearing before the utility vertex with index $u_{\max_I}$ in the decision sequence. Let $C_I^i = \{z \in \mathbb{V} : i \leq z \leq j_I\}$. The conditional expected utility function of equations (2.5.2)-(2.5.5) can be (less intuitively)

written as

$$\bar{u}_i(\boldsymbol{y}_{u_i}) = \sum_{I \in \mathcal{P}_0([m]_{l-1})} \bar{u}_i^I(\boldsymbol{y}_{B_i}) = \sum_{I \in \mathcal{P}_0(\{l,...,m\})} k^{n_I-1} \prod_{s \in I} k_s u_s(\boldsymbol{y}_{P_s}) \sum_{\boldsymbol{y}_{C_I^i} \in \mathcal{Y}_{C_I^i}} P(\boldsymbol{y}_{C_I^i} \mid \boldsymbol{y}_{B_i}),$$
(A.4.7)

where

$$P(\boldsymbol{y}_{C_I^i} \mid \boldsymbol{y}_{B_i}) = \prod_{t \in C_I^i} P(y_t \mid \boldsymbol{y}_{B_t}),$$
(A.4.8)

and $P(\boldsymbol{y}_A \mid \boldsymbol{y}_B) = \mathbb{P}(\boldsymbol{Y}_A = \boldsymbol{y}_A \mid \boldsymbol{Y}_B = \boldsymbol{y}_B)$. The conditional expected utility therefore depends on the power set of the indices of the utility vertices subsequent to Y_i in the decision sequence. We can note that for any $I, J \in \mathcal{P}_0([m]_{l-1})$ such that $\#I = \#J$ and $u_{\max_I} = u_{\max_J}$, $\bar{u}_i^I(\boldsymbol{y}_{B_i})$ and $\bar{u}_i^J(\boldsymbol{y}_{B_i})$ have the same polynomial structure since $C_I^i = C_J^i$. Now for $a = l, \ldots, m$ and $b = l, \ldots, a$, the binomial coefficient $\binom{a-l}{b-l}$ counts the number of elements $I \in \mathcal{P}_0([m]_{l-1})$ having $\#I = b - l + 1$ and including a. Thus r_{iba} in equation (5.6.2) counts the correct number of monomials having a certain degree since $\mathcal{Y}_{C_I(i)} = \times_{t \in C_I^i} \mathcal{Y}_t$. Further note that considering each combination of b and a in the ranges specified above, we consider each element of $\mathcal{P}_0([m]_{l-1})$.

By having a closer look at d_{iba} in equation (5.6.2) it is easy to deduce the corresponding degree of these monomials. The first term of d_{iba}, $(b-l)$, computes the degree associated to the criterion weight k, since $b - l = n_I - 1$ and the second term, $2(b-l+1)$, computes the degree associated to the product between the criterion weights k_s and the utilities $u_s(\boldsymbol{y}_{P_s})$ for $s \in C_I^i$. The last term w_{ia} corresponds to the degree deriving from the probabilistic part of equation (A.4.7), which is equal to the number of non-controlled vertices between Y_i and $Y_{j_{\max_I}}$ (both included) as shown by equation (A.4.8).

Since the set B_i includes the arguments of $\bar{u}_i(\boldsymbol{y}_{B_i})$ and $\boldsymbol{\mathcal{Y}} = \times_{i \in [n]} \mathcal{Y}_i$, equation (5.6.1) guarantees that the dimension of the conditional expected utility vector is $\prod_{t \in B_i} r_t$.

A.4.4 Proof of Theorem 5.6.12

For $i, j, k, l \in \mathbb{V}$ and $s, t \in [m]$, an asymmetry $Y_i = y_i \Rightarrow Y_j = y_j$ implies that any monomials that include terms of the form $p_{ky_k\pi(k)}$, $\theta_{s\pi(s)}$, $p_{ky_k\pi(k)}p_{ly_l\pi(l)}$, $\theta_{t\pi(t)}\theta_{s\pi(s)}$ and $p_{ky_k\pi(k)}\theta_{s\pi(s)}$ entailing both instantiations y_i and y_j are associated to a non possible combination of events, with $y_k \in \mathcal{Y}_k$, $\pi(k) \in \mathcal{Y}_{\Pi(k)}$, $y_l \in \mathcal{Y}_l$, $\pi(l) \in \mathcal{Y}_{\Pi(l)}$, $\pi(t) \in \mathcal{Y}_{P(t)}$ and $\pi(s) \in \mathcal{Y}_{P(s)}$. Thus these monomials have to be set equal to zero.

For $j < t \leq z$, $\bar{u}_t$ has an associated set B_t which includes both i and j and consequently $\prod_{s \in B_t \setminus \{i \cup j\}} r_s$ rows of the vector corresponds to the conditioning on $Y_i = y_i$ and $Y_j = y_j$. Therefore all the monomials in those rows have to be set equal to zero.

For $i < t \leq j$, the index i is in the set B_t, whilst the variable Y_j has been already EUMarginalised. Thus, there are only $\prod_{s \in B_t \setminus \{i\}} r_s$ rows conditional on the event $Y_i = y_i$. In those rows only some of the monomials are associated to the event $Y_j = y_j$. Specifically, the ones implying $Y_j = y_j$ can only be multiplying a term including a $\boldsymbol{\theta}_{xP_x}$ from a utility vertex u_x subsequent to Y_j in the MID DS. We can deduce that there are $\prod_{s=t}^{ja} r_s / r_j$ monomials of degree d_{tba} that include the case $Y_j = y_j$ in such entries of $\bar{\boldsymbol{u}}_t$, for $a \in [m]_{x-1}$ and $b \in [a]_{l-1}$ (using the notation of Theorem 5.6.3).

Lastly, if $t \leq i$, then the set B_t does not include i and j, which have been both EU-Marginalised. Thus monomials including a combination of the events $Y_j = y_j$ and $Y_i = y_i$ appears in each row of $\bar{\boldsymbol{u}}_t$. Just as before, we can deduce that there are $\prod_{s=t}^{ja} r_s / (r_i \cdot r_j)$ monomials of degree d_{tba}, $a \in [m]_{x-1}$, $b \in [a]_{l-1}$, implying the event $Y_i = y_i \wedge Y_j = y_j$.

Appendix B

Graph Theory

B.1 Some Definitions

Definition B.1.1. *A **graph** $\mathcal{G}$ is a pair $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$, where $V(\mathcal{G})$ is a finite set of vertices, or nodes, of $\mathcal{G}$ and $E(\mathcal{G})$ is a subset of $V(\mathcal{G}) \times V(\mathcal{G})$ of ordered distinct pairs of vertices, called* edges, *or arcs, of $\mathcal{G}$.*

Note that our definition does not allow for the presence of multiple edges between the same two vertices. Edges connecting one vertex to itself are also not allowed by the definition above. Throughout the thesis we consider generic graphs whose vertex set is equal to $\{Y_i : i \in [n]\}$, $\{\boldsymbol{Y}_i^T : i \in [n]\}$ or $\{\boldsymbol{r}_i : i \in [m]\}$, among the others. In this appendix we consider the vertex set $\{Y_i : i \in [n]\}$, unless otherwise specified, to define the terminology employed in this thesis. This however straightforwardly applies to any other vertex set.

Let $i, j \in [n]$. For Y_i, $Y_j \in V(\mathcal{G})$, if $(Y_i, Y_j), (Y_j, Y_i) \in E(\mathcal{G})$ we say that there is an *undirected* edge between Y_i and Y_j. The vertices Y_i and Y_j are said to be *neighbours*. The set of the indices of the neighbours of a vertex Y_i is denoted by Ne_i. If $(Y_i, Y_j) \in E(\mathcal{G})$ but $(Y_j, Y_i) \notin E(\mathcal{G})$, we say that there is a *directed* edge *from Y_i to Y_j*. We further say that Y_i is a *parent* of Y_j and that Y_j is a *child* of Y_i. The set of the indices of the parents of Y_i is denoted with Π_i, whilst Ch_i is the set of the indices of its children. We define $Fa_i = \{i\} \cup \Pi_i$ and we call $\{Y_i\} \cup \{Y_j : j \in \Pi_i\}$ the *family* of Y_i. The *boundary* of Y_i is $\{Y_j : j \in Bd_i\}$, where $Bd_i = \Pi_i \cup Ne_i$. If $(Y_i, Y_j) \in E(\mathcal{G})$ and/or $(Y_j, Y_i) \in E(\mathcal{G})$, we say that Y_i and Y_j are joined.

A *path* of length m from Y_{i_1} to Y_{j_m} in a graph $\mathcal{G}$ is a sequence of m distinct edges in $E(\mathcal{G})$

$$((Y_{i_1}, Y_{j_1}), \ldots, (Y_{i_k}, Y_{j_k}), (Y_{i_{k+1}}, Y_{j_{k+1}}), \ldots, (Y_{i_m}, Y_{j_m})),$$

such that $Y_{j_k} = Y_{i_{k+1}}$, $k \in [m-1]$, $i_k, j_k \in [n]$ and we say that the path *ends* in Y_{j_m}. A *cycle* is a path with the additional condition that $Y_{i_1} = Y_{j_n}$. A *rooted path* of length $m+1$ from Y_{i_1} to Y_{j_m} is a sequence comprising of a vertex in $V(\mathcal{G})$ and m distinct edges in $E(\mathcal{G})$ is such that

$$(Y_{i_1}, (Y_{i_1}, Y_{j_1}), \ldots, (Y_{i_k}, Y_{j_k}), (Y_{i_{k+1}}, Y_{j_{k+1}}), \ldots, (Y_{i_m}, Y_{j_m})),$$

where $Y_{j_k} = Y_{i_{k+1}}$, $k \in [m-1]$, $i_k, j_k \in [n]$. Cycles and (rooted) paths including at least one directed edge are called *directed* cycles and directed (rooted) paths respectively.

If in $\mathcal{G}$ there are both a path from Y_i to Y_j and a path from Y_j to Y_i we say that Y_i and Y_j are *connected*. It can be seen that the vertex set of a graph $\mathcal{G}$ can be uniquely partitioned into subsets of vertices such that every element of the subset is connected with all the other elements of that subset. We call such subsets the *strong components* of $\mathcal{G}$. We further let a *trail* from Y_i to Y_j be a path where two consecutive vertices of the associated sequence simply needs to be joined.

Definition B.1.2. *We define the following special graphs:*

- *if all the edges of $\mathcal{G}$ are undirected we say that $\mathcal{G}$ is an* Undirected Graph (UG)*;*

- *if all the edges of $\mathcal{G}$ are directed we say that $\mathcal{G}$ is a* directed graph*;*

- *if a graph $\mathcal{G}$ includes both directed and undirected graphs we say that $\mathcal{G}$ is a* mixed graph*;*

- *the* undirected version *of a generic graph $\mathcal{G}$, denoted by $\mathcal{G}^U$ is the undirected graph obtained by replacing every directed edge with an undirected one;*

- *a graph $\mathcal{G}' = (V(\mathcal{G}'), E(\mathcal{G}'))$ is a* subgraph *of $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$ if $V(\mathcal{G}') \subseteq V(\mathcal{G})$ and $E(\mathcal{G}') \subseteq E(\mathcal{G}) \cap \{V(\mathcal{G}') \times V(\mathcal{G}')\}$. If $E(\mathcal{G}') = E(\mathcal{G}) \cap \{V(\mathcal{G}') \times V(\mathcal{G}')\}$ we say that $\mathcal{G}'$ is the subgraph of $\mathcal{G}$ induced by $V(\mathcal{G}')$;*

- *a graph $\mathcal{G}$ is* complete *if each pair of its vertices is joined;*

- *a* clique *$\{Y_i : i \in C\}$, $C \subseteq [n]$, of a graph $\mathcal{G}$ is such that the subgraph induced by $\{Y_i : i \in C\}$ is a complete maximal subgraph of $\mathcal{G}$, where maximal means that there is no $C \subset C'$ such that the subgraph induced by $\{Y_i : i \in C'\}$ is complete;*

- *a graph $\mathcal{G}$ is* connected *if there is a trail between every pair of vertices in $\mathcal{G}$;*

- *a* Directed Acyclic Graph (DAG) *is a directed graph with no directed cycles;*

- *a* Chain Graph (CG) *is a mixed graph with no directed cycles;*

Note in particular that DAGs and UGs are special cases of CGs. Furthermore the strong components of a DAG all consists of a single vertex.

B.2 Chain Graphs and Directed Acyclic Graphs

We now focus in more detail on CGs. We first note that it is always possible, although not uniquely, to *well-order* the vertices of a CG with an indexing such that if there is a directed edge from Y_i to Y_j, then $i < j$. See for example Cowell et al. [1999] for algorithms that allow for a construction of such an indexing.

For a well-ordered CG $\mathcal{G}$, we say that Y_i is an *ancestor* of Y_j if there is a path from Y_i to Y_j but not a path from Y_j to Y_i. In this case Y_j is also call a *descendant* of Y_i and we denote with De_i the set of the indices of descendants of Y_i. The *ancestral set* of a vertex Y_i is the set including all its ancestors and Y_i itself. We denote with A_i the set of indices of the variables in the ancestral set and we define $A'_i = A_i \setminus \{i\}$ (this is usually referred to as the *non-descendant set*). More generally the ancestral set of $\{Y_i : i \in A\}$, $A \subseteq [n]$, is the union of the set of ancestors of every element in $\{Y_i : i \in A\}$ together with the elements in $\{Y_i : i \in A\}$. The *ancestral graph* of $\{Y_i : i \in A\}$ is the subgraph of $\mathcal{G}$ induced by $\{Y_i : i \in A\}$.

Let A, B and C be three disjoint subsets of $[n]$. We say that $\{Y_i : i \in C\}$ *separates* $\{Y_i : i \in A\}$ from $\{Y_i : i \in B\}$ if every trail from any element in $\{Y_i : i \in A\}$ to any element in $\{Y_i : i \in B\}$ goes through at least one element in $\{Y_i : i \in C\}$. We now define a special type of chain graph.

Definition B.2.1. *Given a CG $\mathcal{G}$, the* moral graph, $\mathcal{G}^m$, *of $\mathcal{G}$ is the undirected graph obtained from $\mathcal{G}$ by the following procedure:*

- *add an undirected edge between any pair of vertices that have children in a common strong component and that are not already joined;*

- *form the undirected version of the obtained graph.*

Let $\mathcal{G}$ now be a well-ordered DAG. Note that for a DAG the moralisation process adds an edge between all pairs of parents of the same children not already joined by a directed edge. The resulting graph from the moralisation of a BN is said to be *decomposable*. We say that a vertex Y_i is the *father* of Y_j and Y_j is its *son*, for $Y_i, Y_j \in \mathcal{G}$, if $(Y_i, Y_j) \in E(\mathcal{G})$ and there is no other directed path from Y_i to Y_j. Note that in a decomposable DAG each vertex either has one or no father. We denote with F_i the index of the father of Y_i, whilst S_i is the set of the indices of its sons. A vertex of a DAG with no children is

called *leaf*, whilst a *root* is a vertex with no parents. We let *Le* be the index set of the leafs of the DAG.

Consider now an undirected graph $\mathcal{G}$ and let σ be a cycle. A *chord* of this cycle is an edge $(Y_i, Y_l) \in \mathbb{E}(\mathcal{G})$ such that the edges $(Y_i, Y_j), (Y_k, Y_l) \in E(\mathcal{G})$ are non consecutive elements of σ. The undirected graph $\mathcal{G}$ is said to be *chordal* if all its cycles of length at least four have a chord. It is possible to show that the notion of chordal graph coincides with the one of decomposable graph (which for the purpose of this thesis we omit). Thus generally we refer to a chordal graph as a decomposable graph: see for more details Cowell et al. [1999]. We also say that a chain graph is decomposable meaning that its undirected version is chordal.

The last concept concerning undirected graphs we define here is the one of decomposition.

Definition B.2.2. *The sets $\{Y_i : i \in A\}$ and $\{Y_i : i \in B\}$, $A, B \subset [n]$ are said to form a* decomposition *of a graph $\mathcal{G}$ with vertex set $\{Y_i : i \in [n]\}$ if*

- $A \cup B = [n]$;

- $\{Y_i : i \in A \cap B\}$ *separates* $\{Y_i : i \in A\}$ *from* $\{Y_i : i \in B\}$;

- *the subgraph of $\mathcal{G}$ induced by $\{Y_i : i \in A \cap B\}$ is complete.*

B.3   Trees

We now deal with a special class of graphs called *trees*.

Definition B.3.1. *A graph $\mathcal{G}$ is a* tree *if it is connected and its undirected version has no cycles.*

Note that in a tree there is a unique trail between any two vertices. A rooted tree is a tree with a designated vertex called *root*. A *leaf* of a tree is a vertex which is joined to at most one other node. A *forest* is a graph having no cycles, i.e. its strong components are all trees.

An important class of trees is defined below.

Definition B.3.2. *Let $\mathcal{G}$ be an undirected graph, $\{Y_j : j \in C_i\}$, $i \in [m]$, its cliques and $\mathcal{C} = \{C_1, \dots, C_m\}$. A* junction tree *of $\mathcal{G}$ has vertex set $\{Y_C : C \in \mathcal{C}\}$ and is such that $C_i \cap C_j \subset C_k$, $i, j, k \in [m]$, for any C_k such that $\{Y_l : l \in C_k\}$ is a member of an edge in the unique path between $\{Y_l : l \in C_i\}$ and $\{Y_l : l \in C_j\}$.*

It is possible to show that if $\mathcal{G}$ is decomposable, then there exists a unique junction tree associated to that graph and its cliques can be ordered so that they exhibit the *running intersection property*, i.e., for all $j \in [m]$, there is an $i < j$ such that $S_j = C_j \cap \cup_{k \in [j-1]} C_k \subseteq C_i$. Note that $\{Y_i : i \in S_j\}$, separates $\{Y_i : i \in C_j\}$ from $\{Y_i : i \in \cup_{k \in [j-1]} C_k\}$ and we therefore call it a *separator*. Let also $\mathcal{S} = \{S_2, \ldots, S_m\}$. Note that by construction $S_j \subset C_i$ for an $i < j$. Cowell et al. [1999] presented several algorithms to construct a junction tree exhibiting the running intersection property.

Another class of trees we deal with is introduced now.

Definition B.3.3. *A directed tree $\mathcal{T}$ is a directed graph with the following two properties:*

- *it has a unique vertex with no parents called root;*

- *all other vertices have exactly one parent.*

It can be noted that a directed tree is also a generic tree according to Definition B.3.1. The vertex set of a directed tree, $V(\mathcal{T})$, is partitioned into the set of leaves $L(\mathcal{T})$, the vertices with no children, and the set of *situations* $S(\mathcal{T}) = V(\mathcal{T}) \setminus L(\mathcal{T})$. A *floret* $\mathcal{F}(v_i)$ of a situation $v_i \in S(\mathcal{T})$ is the subgraph of $\mathcal{T}$ generated by $\{v_i \cup v_{Ch_i}\}$. Note that any directed tree is fully defined by its set of florets.

Appendix C

Bayesian Models

C.1   Discrete Models

C.1.1   Beta-Binomial Model

Let $Y \mid \theta \sim \text{Bin}(n, \theta)$ for $n \in \mathbb{Z}_{\geq 1}$ and $\theta \in [0, 1]$, i.e., given θ, Y is a *Binomial* distribution associated to n independent trials with probability θ of success. Its density function is written as

$$f(y \mid \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y},$$

where $\binom{n}{y}$ is a Binomial coefficient. Recall that $\mathbb{E}(Y \mid \theta) = n\theta$ and $\mathbb{V}(Y \mid \theta) = n\theta(1 - \theta)$. Assume $\theta \sim \text{Be}(p, q)$, $p, q \in \mathbb{Z}_{>0}$, i.e. θ follows a *Beta* distribution whose density can be written as

$$\pi(\theta) = \frac{1}{\text{B}(p, q)} \theta^{p-1} (1 - \theta)^{q-1},$$

where $\text{B}(p, q)$ is the *Beta function* equal to

$$\text{B}(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p + q)}, \tag{C.1.1}$$

with

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} \mathrm{d}x,$$

for $p \in \mathbb{R}_{>0}$. Note that, if $p \in \mathbb{Z}_{\geq 1}$, then $\Gamma(p) = p - 1!$. The expectation and the variance of θ can be seen to be equal to

$$\mathbb{E}(\theta) = \frac{p}{p + q}, \qquad \mathbb{V}(\theta) = \frac{pq}{(p + q)^2 (p + q + 1)}.$$

By applying Bayes theorem in equation (2.1.2) the posterior density of $\theta \mid Y$ is proportional to

$$\theta^{y+p-1}(1-\theta)^{q+n-y-1},$$

and therefore $\theta \mid Y \sim \mathrm{Be}(y+p, q+n-y)$. Consequently beta priors are conjugate with binomial likelihoods.

C.1.2 Dirichlet-Multinomial Model

Let $Z_1, \ldots, Z_n$, $n \in \mathbb{Z}_{\geq 1}$ be independent and identically distributed random variables taking values in $k \in \mathbb{Z}_{\geq 1}$ possible categories with probabilities $\theta_1, \ldots, \theta_k$, $\theta_i \in [0,1]$, $i \in [k]$. Let Y_i, $i \in [k]$, denote the random variable counting the number of observations of the Zs in the i-th category. By construction $n = \sum_{i \in [k]} y_i$. Letting $\boldsymbol{\theta} = (\theta_i)_{i \in [k]}^{\mathrm{T}}$, $\boldsymbol{Y} = (Y_i)_{i \in [k]}^{\mathrm{T}}$, we have that $\boldsymbol{Y} \mid \boldsymbol{\theta}$ is a *Multinomial* distribution, denoted as $\boldsymbol{Y} \mid \boldsymbol{\theta} \sim \mathrm{Multi}(n, \boldsymbol{\theta})$, whose density can be written as

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) = \binom{n}{\boldsymbol{y}} \prod_{i \in [k]} \theta_i^{y_i},$$

where

$$\binom{n}{\boldsymbol{y}} = \frac{n!}{y_1! \cdots y_k!},$$

is a *Multinomial coefficient*. Note that marginally each $Y_i \mid \boldsymbol{\theta}$, $i \in [k]$, behaves as a Binomial distribution with parameters n and θ_i, and $\mathbb{C}(Y_i, Y_j \mid \boldsymbol{\theta}) = -n\theta_i\theta_j$, $i \neq j$, $i, j \in [k]$.

Assume $\boldsymbol{\theta}$ follows a *Dirichlet* distribution of parameter $\boldsymbol{a} = (a_i)_{i \in [k]}^{\mathrm{T}} \in \mathbb{R}_{>0}^k$, $\boldsymbol{\theta} \sim \mathrm{Di}(\boldsymbol{a})$, with density

$$\pi(\boldsymbol{\theta}) = \frac{1}{\mathrm{B}(\boldsymbol{a})} \prod_{i \in [k]} \theta_i^{a_i - 1},$$

where $\mathrm{B}(\boldsymbol{a})$ is a generalisation of the beta function in equation (C.1.1) equal to

$$\mathrm{B}(\boldsymbol{a}) = \frac{\prod_{i \in [k]} \Gamma(a_i)}{\Gamma\left(\sum_{i \in [k]} a_i\right)}.$$

Note that for any $I \subset [k]$, $\boldsymbol{\theta}_I \sim \mathrm{Di}(\boldsymbol{a}_I)$, where $\boldsymbol{\theta}_i = (\theta_i)_{i \in I}^{\mathrm{T}}$ and $\boldsymbol{a}_I = (a_i)_{i \in I}^{\mathrm{T}}$, and that $\theta_i \sim \mathrm{Be}(a_i, |\boldsymbol{a}| - a_i)$, $i \in [k]$, where $|\boldsymbol{a}| = \sum_{i \in [k]} a_i$. The first two moments have a closed form which can be easily deduced to be, for $i, j \in [k]$, $i \neq j$,

$$\mathbb{E}(\theta_i) = \frac{a_i}{|\boldsymbol{a}|}, \quad \mathbb{V}(\theta_i) = \frac{a_i(|\boldsymbol{a}| - a_i)}{a^2(|\boldsymbol{a}| + 1)}, \quad \mathbb{C}(\theta_i, \theta_j) = -\frac{a_i a_j}{a^2(|\boldsymbol{a}| + 1)}.$$

Order	moment
1	μ
2	$\psi + \mu^2$
3	$\mu^3 + 3\mu\psi$
4	$\mu^4 + 6\mu^2\psi + 3\psi^2$
5	$\mu^5 + 10\mu^3\psi + 15\mu\psi^2$
6	$\mu^6 + 15\mu^4\psi + 45\mu^2\psi^2 + 15\psi^3$

TABLE C.1: First six moments of a Normal distribution with mean μ and variance ψ.

We can note that the Dirichlet family is conjugate with the Multinomial one by simply applying Bayes theorem in equation (2.1.2). It holds that

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \prod_{i \in [k]} \theta_i^{y_i + a_i - 1},$$

and therefore $\boldsymbol{\theta} \mid \boldsymbol{Y} \sim \mathrm{Di}(\boldsymbol{a} + \boldsymbol{y})$.

C.2 Continuous Models

C.2.1 Normal Inverse Gamma Models

Let $Y \mid \mu, \psi \sim \mathrm{N}(\mu, \psi)$, where $\mu \in \mathbb{R}$ and $\psi \in \mathbb{R}_{>0}$, i.e. Y follows a *Normal* (or *Gaussian*) distribution given the parameters μ and ψ.[1] Recall that $\mathbb{E}(Y \mid \mu, \psi) = \mu$, $\mathbb{V}(Y \mid \mu, \psi) = \psi$ and that the density is written as

$$f(y \mid \mu, \psi) = \frac{1}{\sqrt{2\pi\psi}} \exp\left(-\frac{1}{2\psi}(y - \mu)^2\right). \tag{C.2.1}$$

The distribution of a Normal random variable is fully characterised by its first two moments since every higher order moment is a function of these. We summarise the first moments of a Normal distribution in Table C.1.

Suppose the prior distribution over (μ, ψ) has density

$$\pi(\mu, \psi) = \frac{\sqrt{r}}{\sqrt{2\pi\psi}} \frac{(a/2)^{b/2}}{\Gamma(b/2)} (\psi)^{-b/2-1} \exp\left(-\frac{a + r(\mu - m)^2}{2\psi}\right), \tag{C.2.2}$$

---

[1]We choose to parametrise the variance with ψ instead of the more common ψ^2 in order to highlight the degree of the moments required for our computations.

for $m \in \mathbb{R}$ and $a, b, r \in \mathbb{R}_{>0}$. Note that the joint density $\pi(\mu, \psi)$ can be written as the product of $\pi(\mu \mid \psi)$ and $\pi(\psi)$, where

$$\pi(\mu \mid \psi) = \frac{\sqrt{r}}{\sqrt{2\pi\psi}} \exp\left(-\frac{r(\mu - m)^2}{2\psi}\right),$$

$$\pi(\psi) = \frac{(a/2)^{b/2}}{\Gamma(b/2)} (\psi)^{-a/2-1} \exp\left(-\frac{a}{2\psi}\right).$$

Therefore $\mu \mid \psi \sim \mathrm{N}(m, \psi/r)$ and $\psi \sim \mathrm{IG}(a, b)$, meaning that ψ follows an *Inverse Gamma* distribution with parameters a and b. We say that a distribution having the density in equation (C.2.2) is a *Normal Inverse Gamma* with parameters m, r, a and b, i.e. $(\mu, \psi) \sim \mathrm{NIG}(m, r, a, b)$. We study in more details the features of this distribution in the following section.

Now the posterior density $(\mu, \psi) \mid y$ can be deduced by multiplying equations (C.2.1) and (C.2.2) and is proportional to

$$\frac{1}{\sqrt{2\pi\psi}} \exp\left(-(1+r)\left(\mu - \frac{y+rm}{1+r}\right)^2 / 2\psi\right) (\psi)^{-b-\frac{3}{2}} \exp\left(-\frac{1}{2\psi}\left(a + \frac{r}{1+r}(y-m)^2\right)\right).$$

(C.2.3)

Equation (C.2.3) is proportional to the density of a $\mathrm{NIG}(m', r', a', b')$, where

$$\begin{aligned}
m' &= \tfrac{y+rm}{1+r}, & r' &= 1 + r, \\
b' &= b + \tfrac{1}{2}, & a' &= a + \tfrac{1}{2}\tfrac{r}{1+r}(y-m)^2.
\end{aligned}$$

Therefore an NIG distribution is conjugate with a Normal likelihood.

C.2.2 Bayesian Normal Linear Models

We now consider a *Normal linear model* defined by

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y} = (y_i)_{i \in [n]}^{\mathrm{T}}$, $n \in \mathbb{Z}_{\geq 1}$, is the vector of observations from $\boldsymbol{Y}$, X is a $n \times p$ matrix of known values of covariates, $p \in \mathbb{Z}_{\geq 1}$, $\boldsymbol{\beta} = (\beta_i)_{i \in [p]}^{\mathrm{T}} \in \mathbb{R}^p$ is a vector of parameters and $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i \in [n]}^{\mathrm{T}}$ is a vector of random errors, where $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed Normal random variables with mean 0 and variance $\psi \in \mathbb{R}_{>0}$. Therefore $\boldsymbol{Y} \mid \boldsymbol{\beta}, \psi \sim \mathrm{N}(X\boldsymbol{\beta}, \psi I_n)$, where I_n is the identity matrix of dimension $n \times n$. The density of this *multivariate Normal* is equal to

$$f(\boldsymbol{y} \mid \boldsymbol{\beta}, \psi) = (2\pi\psi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\psi}(\boldsymbol{y} - X\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{y} - X\boldsymbol{\beta})\right).$$

Assume a *multivariate NIG* prior is given to $(\boldsymbol{\beta}, \psi)$ with parameters $(\boldsymbol{m}, V, a, b)$, whose density can be written as

$$\pi(\boldsymbol{\beta}, \psi) = \left(\frac{a}{2}\right)^{\frac{b}{2}} (\psi)^{-\frac{b+p+2}{2}} \frac{1}{(2\pi)^{\frac{p}{2}} \det(V)^{\frac{1}{2}} \Gamma(\frac{b}{2})} \exp\left(-\frac{(\boldsymbol{\beta}-\boldsymbol{m})^{\mathrm{T}} V^{-1} (\boldsymbol{\beta}-\boldsymbol{m}) + a}{2\psi^2}\right),$$

(C.2.4)

where V is a $p \times p$ positive semidefinite symmetric matrix, $\boldsymbol{m} \in \mathbb{R}^n$, $a, b \in \mathbb{R}_{>0}$ and $\det(V)$ is the determinant of V. Note that again this is the product of a Normal distribution over $\boldsymbol{\beta} \mid \psi$ and an Inverse Gamma over ψ. The marginal moments of $\boldsymbol{\beta}$ and ψ can be easily deduced and are equal to

$$\mathbb{E}(\boldsymbol{\beta}) = \boldsymbol{m}, \qquad \mathbb{V}(\boldsymbol{\beta}) = \frac{a}{b-2} V,$$
$$\mathbb{E}(\psi) = \frac{a}{b-2}, \qquad \mathbb{V}(\psi) = \frac{2a^2}{(b-2)^2(b-4)}.$$

Note that $\boldsymbol{\beta}$ marginally follows a *multivariate T* distribution with b degrees of freedom and parameters $\boldsymbol{m}$ and aV, denoted as $\boldsymbol{\beta} \sim \mathrm{T}_b(\boldsymbol{m}, aV)$, whose density is equal to

$$\pi(\boldsymbol{\beta}) = \frac{a^{b/2} \Gamma((b+p)/2)}{\det(V)^{1/2} \pi^{p/2} \Gamma(b/2)} (a + (\boldsymbol{\beta}-\boldsymbol{m})^{\mathrm{T}} V^{-1} (\boldsymbol{\beta}-\boldsymbol{m}))^{(b+p)/2}.$$

Using again Bayes theorem in equation (2.1.2) it can be shown that the NIG density in equation (C.2.4) is conjugate for the Normal linear model. Its posterior is then $\mathrm{NIG}(\boldsymbol{m}', V', a', b')$, where

$$V' = \left(V^{-1} + X^{\mathrm{T}} X\right)^{-1}, \qquad \boldsymbol{m}' = \left(V^{-1} + X^{\mathrm{T}} X\right)^{-1} \left(V^{-1} \boldsymbol{m} + X^{\mathrm{T}} \boldsymbol{y}\right),$$
$$b' = b + n, \qquad a' = a + \boldsymbol{m}^{\mathrm{T}} V^{-1} \boldsymbol{m} + \boldsymbol{y}^{\mathrm{T}} \boldsymbol{y} - \boldsymbol{m}'^{\mathrm{T}} V'^{-1} \boldsymbol{m}'.$$

C.2.3 Multivariate Normal Models

We now consider a generalisation of the previous sections and we now sample from $\boldsymbol{Y}^{\mathrm{T}} = (\boldsymbol{Y}_i^{\mathrm{T}})_{i \in [n]}$, where each $\boldsymbol{Y}_i \mid \boldsymbol{\mu}, \Psi \sim \mathrm{N}(\boldsymbol{\mu}, \Psi)$, $i \in [n]$, and $\perp\!\!\!\perp_{i \in [n]} \boldsymbol{Y}_i \mid \boldsymbol{\mu}, \Psi$. We let $\boldsymbol{\mu} \in \mathbb{R}^m$ and Ψ be a $m \times m$ symmetric positive semidefinite matrix. In this setting

$$f(\boldsymbol{y} \mid \boldsymbol{\mu}, \Psi) = (2\pi)^{-mn/2} \det(\Psi)^{-n/2} \exp\left(-\frac{\sum_{i \in [n]} (\boldsymbol{y}_i - \boldsymbol{\mu})^{\mathrm{T}} \Psi^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu})}{2}\right).$$

Note that the difficulty here is that it needs to be given a prior distribution to the covariance matrix Ψ. The approach to elicit such prior is similar to the ones above and factorises the density $\pi(\boldsymbol{\mu}, \Psi)$ as $\pi(\boldsymbol{\mu} \mid \Psi)\pi(\Psi)$. The distribution of $\boldsymbol{\mu} \mid \Psi$ is assumed to be Normal with mean $\boldsymbol{m}$ and covariance $c^{-1}\Psi$, $c \in \mathbb{R}_{>0}$, whilst Ψ is assumed to follow the so called *Inverse Wishart* distribution, denoted as IW. If $\Psi \sim \mathrm{IW}(A, d)$, where A is

positive semidefinite symmetric $m \times m$ matrix and $d \in \mathbb{R}_{>m-1}$, then the density is equal to

$$\pi(\Psi) = k^{-1} \det(A)^{d/2} \det(\Psi)^{-(d+m+1)/2} \exp\left(-\frac{\text{tr}(\Psi^{-1}A)}{2}\right),$$

where tr denotes the trace of a matrix and

$$k = 2^{dm/2} \pi^{m(m-1)/4} \prod_{i \in [m]} \Gamma((d+1-i)/2).$$

The joint distribution of $(\boldsymbol{\mu}, \Psi)$ so defined follows a *Normal Inverse Wishart* distribution, denoted as NIW. Therefore, if we let $(\boldsymbol{\mu}, \Psi) \sim \text{NIW}(\boldsymbol{m}, c, A, d)$, its density is equal to

$$\pi(\boldsymbol{\mu}, \Psi) = \frac{\det(A)^{d/2} \det(\Psi)^{-(d+m+2)/2}}{(2\pi)^{m/2} k} \exp\left(-\frac{c(\boldsymbol{\mu}-\boldsymbol{m})^{\text{T}} \Psi^{-1}(\boldsymbol{\mu}-\boldsymbol{m}) + \text{tr}(\Psi^{-1}A)}{2}\right).$$

The moments of the Normal distribution can be straightforwardly deduced, whilst for the Inverse Wishart the computation of moments is more complicated. We report here the mean of Ψ, whilst its second moment can be found in Siskind [1972]. Specifically

$$\mathbb{E}(\boldsymbol{\mu}) = \boldsymbol{m}, \qquad \mathbb{V}(\boldsymbol{\mu}) = c^{-1}\Psi, \qquad \mathbb{E}(\Psi) = (d-m-1)^{-1}A.$$

Note that also just as in the Normal linear model, $\boldsymbol{\mu}$ marginally follows a multivariate T distribution with $d+1-m$ degrees of freedom and parameters $(\boldsymbol{m}, c^{-1}A)$.

It can be shown that the NIG distribution is conjugate with the multivariate Normal likelihood and the posterior has parameters $(\boldsymbol{m}', c', A', d')$, where

$$\boldsymbol{m}' = (c+n)^{-1}(c\boldsymbol{m}+n\bar{\boldsymbol{y}}), \qquad\qquad c' = c+n,$$
$$A' = A + nS + cn(c+n)^{-1}(\boldsymbol{m}-\bar{\boldsymbol{y}})^{\text{T}}(\boldsymbol{m}-\bar{\boldsymbol{y}}), \qquad d' = d+n,$$

with $\bar{\boldsymbol{y}} = \sum_{i \in [n]} \boldsymbol{y}_i$ and $S = n^{-1} \sum_{i \in [n]} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^{\text{T}}$.

Appendix D

Polynomial Algebra

In this appendix we introduce the concepts of polynomial algebra that we use in the main body of the thesis [see e.g. Cox et al., 2007, for more details]. We start by defining a monomial.

Definition D.0.1. *A **monomial** with indeterminates $y_1, \ldots, y_n$, $n \in \mathbb{Z}_{\geq 1}$, is*

$$\boldsymbol{y}^{\boldsymbol{\alpha}} = y_1^{\alpha_1} \cdots y_n^{\alpha_n},$$

where $\boldsymbol{y} = y_1 \cdots y_n$ and $\boldsymbol{\alpha} = (\alpha_i)_{i \in [n]}^{\mathrm{T}} \in \mathbb{Z}_{\geq 0}^n$. The degree *of $\boldsymbol{y}^{\boldsymbol{\alpha}}$ is $|\boldsymbol{\alpha}| = \sum_{i \in [n]} \alpha_i$, whilst $\boldsymbol{\alpha}$ is its* multi-degree.

A polynomial is a finite linear combination of monomials with coefficients taking values in a field k.[1] For the purpose of this thesis, we always assume this field to coincide with the real numbers $\mathbb{R}$.

Definition D.0.2. *A **polynomial** $p(\boldsymbol{y})$ with indeterminates $y_1, \ldots, y_n$, and coefficients $c_{\boldsymbol{\alpha}}$ in a field k is defined as*

$$p(\boldsymbol{y}) = \sum_{\boldsymbol{\alpha} \in A} c_{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\alpha}},$$

where $A \subset \mathbb{Z}_{\geq 0}^n$.

Definition D.0.3. *In the notation of Definition D.0.2, a* term *of $p(\boldsymbol{y})$ is $c_{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\alpha}}$ if $c_{\boldsymbol{\alpha}} \neq 0$, and the* degree *of $p(\boldsymbol{y})$ is the maximum $|\boldsymbol{\alpha}|$ such that the coefficient $c_{\boldsymbol{\alpha}}$ is non zero.*

Definition D.0.4. *We say that a polynomial $p(\boldsymbol{y})$ is* homogeneous *if all its terms $c_{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\alpha}}$ have the same degree. Non homogeneous polynomials are called* inhomogeneous*. An homogeneous polynomial such that, for every term $c_{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\alpha}}$, α_i is either 1 or 0 is called* square-free *or* multilinear.

---

[1]A field is a set where addition, subtraction, multiplication and division have the standard properties of operations. A formal definition can be found in most algebra textbooks.

Polynomials with one indeterminate have one term only of a particular degree, whilst polynomials with more than one indeterminate can have many terms with the same degree. In many contexts it is important to order any two terms of a polynomial. We therefore now discuss the notion of monomial ordering. Denote an ordering (i.e. a binary relation) with $>$. First note that there is a one to one correspondence between a monomial $\boldsymbol{y}^{\boldsymbol{\alpha}}$ and its exponent $\boldsymbol{\alpha}$. Therefore any ordering on $\mathbb{Z}_{\geq 0}^n$ induces an ordering on the monomials. A basic requirement that an ordering on monomials needs to entertain is to be *total*.

Definition D.0.5. *An order $>$ on $\mathbb{Z}_{\geq 0}^n$ is total if, for any $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{Z}_{\geq 0}^n$, exactly one of the three statements $\boldsymbol{\alpha} > \boldsymbol{\beta}$, $\boldsymbol{\beta} > \boldsymbol{\alpha}$, $\boldsymbol{\beta} = \boldsymbol{\alpha}$, is true.*

In general, an order on the monomials requires a few more conditions.

Definition D.0.6. *A monomial ordering $>$ on $\mathbb{Z}_{\geq 0}^n$ is a binary relation satisfying:*

- *$>$ is a total ordering on $\mathbb{Z}_{\geq 0}^n$;*

- *for $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{Z}_{\geq 0}^n$,if $\boldsymbol{\alpha} > \boldsymbol{\beta}$ then $\boldsymbol{\alpha} + \boldsymbol{\gamma} > \boldsymbol{\beta} + \boldsymbol{\gamma}$;*

- *every non empty subset of $\mathbb{Z}_{\geq 0}^n$ has a smallest element with respect to $>$.*

Two examples of monomial orderings are the **lexicographical order** and the **reverse lexicographical order**.

Definition D.0.7. *Let $\boldsymbol{\alpha} = (\alpha_i)_{i \in [n]}^{\mathrm{T}}$ and $\boldsymbol{\beta} = (\beta_i)_{i \in [n]}^{\mathrm{T}}$, with $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{Z}_{\geq 0}^n$. We say $\boldsymbol{\alpha} >_{lex} \boldsymbol{\beta}$ with respect to a lexicographic order if in the difference $\boldsymbol{\alpha} - \boldsymbol{\beta}$ the left-most non zero entry is positive. A reverse lexicographic order, $>_{revlex}$, is such that if $\boldsymbol{\alpha} >_{revlex} \boldsymbol{\beta}$ then in the difference $\boldsymbol{\alpha} - \boldsymbol{\beta}$ the right-most non zero entry is positive.*

Lastly, we show that powers of polynomials can be computed applying the *Multinomial theorem* [see e.g. page 336 of Cox et al., 2007].

Theorem D.0.8. *Let $m, n \in \mathbb{Z}_{\geq 1}$, $\boldsymbol{\alpha} \in \mathbb{Z}_{\geq 0}^n$ and $\boldsymbol{y} = y_1 \cdots, y_n$, where y_i is an indeterminate, $i \in [n]$. Then*

$$(y_1 + \cdots + y_n)^m = \sum_{|\boldsymbol{\alpha}| = m} \binom{m}{\boldsymbol{\alpha}} \boldsymbol{y}^{\boldsymbol{\alpha}},$$

where

$$\binom{m}{\boldsymbol{\alpha}} = \frac{m!}{\alpha_1! \cdots \alpha_n!},$$

is the Multinomial coefficient.

Appendix E

Computer Code

Computation of the highest index in each parent set of a utility node.

Inputs: `PiU::table`, parent sets of utility nodes; `m::integer`, number of utility nodes.

Output: `J::list`.

```
CompJ := proc(PiU,m) local i,j:
for j  to m do J[j] := max(PiU[J]) end do:
return convert(J,list):
end proc:
```

Computation of the indices of the argument of the expected utility at step i.

Inputs: `PiU::table`; `PiV::table`, parent sets of random nodes; `i::integer`;
`n::integer`, number of random nodes; `J::list`.

Output: `Bi[i]::set`.

```
CompBi := proc(PiU,PiV,i,n,J) local Bi, part, j:
Bi[i], part := {},{}:
for j from i to n do
  part := part union {j}:
  if member(j,V) then Bi[i] := Bi[i] union PiV[j] end if:
  if member(j,J,'l') then Bi[i] := Bi[i] union PiU[l] end if:
end do:
Bi[i] := Bi[i] minus part:
return Bi[i]:
end proc:
```

Initialisation of an MID.

Inputs: `p::table`, probability vectors; `psi::table`, utility vectors;

```
PiV::table; PiU::table; n::integer; m::integer.
```

Outputs: `J::list`; `Bi::table`; `u::table`, expected utility vectors.

```
Initialise := proc(p, psi, PiV, PiU, n, m) local J, i, Bi, u:
J := CompJ(PiU, m):
for i to n do  Bi[i] := CompBi(PiU, PiV, i, n, J) end do:
Bi[n+1], u[n+1] := {}, []:
return J, Bi, u:
end proc:
```

EUDuplication of a utility vector and an expected utility vector.

Inputs: `u::table`; `psi::table`; `j::integer`; `PiV::table`; `PiU::table`; `r::table`, size of the decision and sample spaces; `Bi::table`; `J::list`.

Outputs: `utemp::list`, EUDuplicated version of `u`; `psitemp::list`, EUDuplicated version of `psi`.

```
EUDuplicationPsi := proc(u, psi, j, PiV, PiU, r, Bi, J)
local i, uprime, psiprime, psitemp, utemp, x, sx, y, h, z:
i, uprime, psiprime:= max(PiU[j]), [], []:
psitemp, utemp:= psi[j], u[i+1]:
for x from max(Bi[i+1], PiU[j]) by -1 to 1 do
  if member(x, (PiU[j] union Bi[i+1]) minus (PiU[j] intersect Bi[i+1]))
    then sx := 1:
    for y from x+1 to max(Bi[i+1], PiU[j]) do
      if member(y, Bi[i+1] union PiU[j]) then sx := sx*r[y] end if
    end do:
    if member(x, Bi[i+1]) then for l to Size(psitemp)[2]/sx do
      for z to r[x] do
      psiprime := [op(psiprime), op(convert(psitemp, list)
      [(l-1)*sx+1 .. l*sx])]
      end do end do:
      psitemp, psiprime := psiprime, []:
    elif member(x, PiU[j]) then for l to Size(utemp)[2]/sx do
      for z to r[x] do
      uprime := [op(uprime),op(convert(utemp, list)[(l-1)*sx+1..l*sx])]
      end do end do:
      utemp,  uprime := uprime, []:
end if end if end do:
return utemp, psitemp:
end proc:
```

EuMultiSum between an expected utility vector and a utility vector.

Inputs: u::table; psi::table; j::integer; PiV::table; PiU::table; r::table; Bi::table; J::list.

Outputs: ut::list, expected utility vector after an EUMultiSum.

```
EUMultiSum := proc(u, psi, j, PiV, PiU, r, Bi, J)
local i, uprime, psiprime, ut;
i := max(PiU[j]);
uprime, psiprime:=EUDuplicationPsi(u, psi, j, PiV, PiU, r, Bi, J);
uprime := convert(uprime, list);
if Size(uprime)[1] = 0 then  ut := k[j]*~psiprime
else ut := h*~k[j]*~psiprime*~uprime+~uprime+~k[j]*~psiprime end if;
return ut:
end proc:
```

EUDuplication of a probability vector and an expected utility vector.

Inputs: u::table; p::table; i::integer; PiV::table; PiU::table; r::table; Bi::table; J::list.

Outputs: utemp::list, EUDuplicated version of u; ptemp::list, EUDuplicated version of p.

```
EUDuplicationP := proc(u, p, i, PiV, PiU, r, Bi, J)
 local uprime, pprime, ptemp, utemp, x, sx, y, l, z:
uprime, pprime, ptemp, utemp   := [], [],  p[i], u[i+1]:
Uni:= Bi[i+1] union PiV[i] union PiU[j];
if member(i, J) then member(i, J, 'j');
  for x from max(Uni) by -1 to 1 do
    if member(x, Uni minus ((Bi[i+1] union PiU[j]) intersect
    (PiV[i] union {i}))) then
      sx := 1;
      for y from x+1 to max(Uni) do
        if member(y, Uni) then  sx := sx*r[y] end if
      end do;
      if member(x, Bi[i+1] union PiU[j]) then
        for l to Size(ptemp)[2]/sx do for z to r[x] do
            pprime := [op(pprime), op(convert(ptemp, list)
            [(l-1)*sx+1 .. l*sx])]
        end do end do;
        ptemp,  pprime := pprime, [];
      elif member(x, PiV[i]) then
```

```
        for l to Size(utemp)[2]/sx do for z to r[x] do
            uprime := [op(uprime), op(convert(utemp, list)
            [(l-1)*sx+1 .. l*sx])]
        end do end do;
        utemp, uprime := uprime, []:
end if end if end do;
else for x from max(Bi[i+1], PiV[i]) by -1 to 1 do
    if member(x, (Bi[i+1] union PiV[i]) minus (Bi[i+1] intersect
    (PiV[i] union {i}))) then
      sx := 1;
      for y from x+1 to max(Bi[i+1], PiV[i]) do
        if member(y, Bi[i+1] union PiV[i]) then sx := sx*r[y] end if
      end do;
      if member(x, Bi[i+1]) then
        for l to Size(ptemp)[2]/sx do for z to r[x] do
            pprime := [op(pprime), op(convert(ptemp, list)
            [(l-1)*sx+1 .. l*sx])]
        end do end do;
        ptemp, pprime := psiprime, [];
      elif member(x, PiV[i]) then for l to Size(utemp)[2]/sx do
       for z to r[x] do
            uprime := [op(uprime), op(convert(utemp, list)
            [(l-1)*sx+1 .. l*sx])]
        end do end do;
        utemp, uprime := uprime, []:
end if end if end do end if:
return utemp, ptemp:
end proc:
```

EUMarginalisation over a sample space.

Inputs: u::table; p::table; i::integer; PiV::table; PiU::table; r::table; Bi::table; J::list.

Outputs: ut::list, expected utility vector after EUMarginalisation.

```
EUMarginalisation := proc(u, p, i, PiV, PiU, r, Bi, J)
local uprime, pprime, rows, cols, l, one, zero, row, Iprime, ut;
uprime, pprime := EUDuplicationP(u, p, i, PiV, PiU, r, Bi, J);
pprime, rows, cols := convert(pprime, list), 1,  Size(pprime)[2];
for l to i-1 do if member(l, PiV[i]) then
```

```
   rows := rows*r[l] end if end do;
one := convert(Vector(r[i], 1), list):
zero := convert(Vector(cols, 0), list):
row, Iprime := [op(zero), op(one)],  [op(one)]:
for l to rows-1 do Iprime := [op(Iprime), op(row)] end do;
Iprime := Matrix(rows, cols, [op(Iprime), op(one)]);
ut := convert(Multiply(Iprime, Vector[column](uprime*~pprime)), list);
return ut:
end proc:
```

Identification of an optimal policy (at random).

Inputs: `r::table`; `i::integer`, index of the decision variable, `t::integer`, number of random draws.

Outputs: `maxi::vector`, optimal decisions.

```
Maximise := proc(r, i, t) local maxi, l:
maxi := Vector(t, 0):
for l to t do
  maxi[l] := RandomTools[Generate](integer(range = 1 .. r[i])) end do:
return maxi:
end proc:
```

EUMaximisation over a decision space.

Inputs: `u::table`; `i::integer`; `r::table`.

Outputs: `u[i]::list`, expected utility vector after EUMaximisation.

```
EUMaximisation := proc(u, i, r) local opt, Istar, j, l, zero;
opt := Maximise(r, i, Size(u[i+1])[2]/r[i]);
zero := convert(Vector(Size(u[i+1])[2], 0), list);
Istar := [];
for j to Size(u[i+1])[2]/r[i] do for l to r[i] do
    if opt[j] = l then  Istar := [op(Istar), 1]
    else Istar := [op(Istar), 0] end if
 end do;
  if j < Size(u[i+1])[2]/r[i] then
     Istar := [op(Istar), op(zero)] end if end do;
Istar := Matrix(Size(u[i+1])[2]/r[i], Size(u[i+1])[2], Istar);
u[i] := convert(Multiply(Istar, Vector[column](u[i+1])), list);
return u[i]:
end proc:
```

Symbolic evaluation algorithm for an MID.

Inputs: `p::table`; `psi::table`; `PiV::table`; `PiU::table`; `n::integer`; `m::integer`; `De::set`, index set of the decision variables; `V::set`, index set of the random variables; `r::table`.

Output: `eu::table`, expected utility vectors.

```
SymbolicExpectedUtility := proc(p, psi, PiV, PiU, n, m, De, V, r)
local J, Bi, utemp, i, j, ceu,best;
with(LinearAlgebra): with(ArrayTools):
J, Bi, eu := Initialise(p, psi, PiV, PiU, n, m);
j := m;
for i from n by -1 to 1 do if j = 0 then if member(i, De) then
      eu[i] := EUMaximisation(eu, i, r)
    else eu[i] := EUMarginalisation(eu, p, i, PiV, PiU, r, Bi, J)
    end if;
  else if J[j] = i then  if member(i, De) then
        utemp[i+1] := EUMultiSum(eu, psi, j, PiV, PiU, r, Bi, J);
        eu[i] := EUMaximisation(utemp, i, r)
      else
        utemp[i+1] := EUMultiSum(eu, psi, j, PiV, PiU, r, Bi, J);
        eu[i] := EUMarginalisation(utemp, p, i, PiV, PiU, r, Bi, J)
      end if;
      j := j-1
    else if member(i, De) then eu[i] := EUMaximisation(eu, i, r)
      else eu[i] := EUMarginalisation(eu, p, i, PiV, PiU, r, Bi, J)
end if end if end if end do;
return eu:
end proc:
```

Consider the MID in Figure 2.18 with $n = 6$ variables (decision or random nodes) and $m = 3$ utility nodes.

Definition of the MID number of variables and utility nodes.

```
n := 6: m := 3:
```

`V`, contains the indices of random nodes and `De` those of the decision nodes.

```
V := {2, 3, 5, 6}: De := {1, 4}:
```

Conditional probabilities.

```
p[6] := [p6111, p6011, p6101, p6001, p6110, p6010, p6100, p6000]:
p[5] := [p5111, p5011, p5101, p5001, p5110, p5010, p5100, p5000]:
p[3] := [p3111, p3011, p3101, p3001, p3110, p3010, p3100, p3000]:
```

```
p[2] := [p211, p201, p210, p200]:
```
Utility parameters.
```
psi[1] := [psi11, psi10]:
psi[2] := [psi21, psi20]:
psi[3] := [psi311, psi301, psi310, psi300]:
```
Parents of random nodes.
```
PiV[2] := {1}:  PiV[3] := {1, 2}:  PiV[5] := {3, 4}:  PiV[6] := {4, 5}:
```
Parents of utility nodes.
```
PiU[1] := {3}: PiU[2] := {5}: PiU[3] := {4, 6}:
```
Number of levels of the variables.
```
r[1] := 2: r[2] := 2: r[3] := 2: r[4] := 2: r[5] := 2: r[6] := 2:
```
Computation of the expected utility vectors.
```
eu := SymbolicExpectedUtility(p, psi, PiV, PiU, n, m, De, V, r):
```

Example of the output of `SymbolicExpectedUtility`, `eu[1]`:

```
[((k[1]*psi11+h*k[1]*psi11*((k[2]*psi21+h*k[2]*psi21*
(p6010*psi300*k[3]+p6110*psi310*k[3])+k[3]*psi300*p6010
+k[3]*psi310*p6110)*p5101+(k[2]*psi20+h*k[2]*psi20*
(p6000*psi300*k[3]+p6100*psi310*k[3])+k[3]*psi300*p6000
+k[3]*psi310*p6100)*p5001)+(k[2]*psi21+h*k[2]*psi21*
(p6010*psi300*k[3]+p6110*psi310*k[3])+k[3]*psi300*p6010
+k[3]*psi310*p6110)*p5101+(k[2]*psi20+h*k[2]*psi20*
(p6000*psi300*k[3]+p6100*psi310*k[3])+k[3]*psi300*p6000
+k[3]*psi310*p6100)*p5001)*p3110+(k[1]*psi10+h*k[1]*psi10*
((k[2]*psi21+h*k[2]*psi21*(p6011*psi301*k[3]
+p6111*psi311*k[3])+k[3]*psi301*p6011+k[3]*psi311*p6111)*p5110+
(k[2]*psi20+h*k[2]*psi20*(p6001*psi301*k[3]+p6101*psi311*k[3])
+k[3]*psi301*p6001+k[3]*psi311*p6101)*p5010)+
(k[2]*psi21+h*k[2]*psi21*(p6011*psi301*k[3]+p6111*psi311*k[3])
+k[3]*psi301*p6011+k[3]*psi311*p6111)*p5110+
(k[2]*psi20+h*k[2]*psi20*(p6001*psi301*k[3]+p6101*psi311*k[3])
+k[3]*psi301*p6001+k[3]*psi311*p6101)*p5010)*p3010)*p210+((k[1]*
psi11+h*k[1]*psi11*((k[2]*psi21+h*k[2]*psi21*(p6010*psi300*
k[3]+p6110*psi310*k[3]+k[3]*psi300*p6010+k[3]*psi310*p6110)*p5101
+(k[2]*psi20+h*k[2]*psi20*(p6000*psi300*k[3]+p6100*psi310*k[3])
+k[3]*psi300*p6000+k[3]*psi310*p6100)*p5001)+
(k[2]*psi21+h*k[2]*psi21*(p6010*psi300*k[3]+p6110*psi310*k[3])
+k[3]*psi300*p6010+k[3]*psi310*p6110)*p5101+
```

```
(k[2]*psi20+h*k[2]*psi20*(p6000*psi300*k[3]+p6100*psi310*k[3])
+k[3]*psi300*p6000+k[3]*psi310*p6100)*p5001)*p3100+(k[1]*psi10
+h*k[1]*psi10*((k[2]*psi21+h*k[2]*psi21*(p6011*psi301*k[3]
+p6111*psi311*k[3])+k[3]*psi301*p6011+k[3]*psi311*p6111)*p5110+
(k[2]*psi20+h*k[2]*psi20*(p6001*psi301*k[3]+p6101*psi311*k[3])
+k[3]*psi301*p6001+k[3]*psi311*p6101)*p5010)+
(k[2]*psi21+h*k[2]*psi21*(p6011*psi301*k[3]+p6111*psi311*k[3])
+k[3]*psi301*p6011+k[3]*psi311*p6111)*p5110+
(k[2]*psi20+h*k[2]*psi20*(p6001*psi301*k[3]+p6101*psi311*k[3])
+k[3]*psi301*p6001+k[3]*psi311*p6101)*p5010)*p3000)*p200]
```

Bibliography

O. O. Aalen, K. Røysland, J. M. Gran, and B. Ledergerber. Causality, mediation and time: a dynamic viewpoint. *Journal of the Royal Statistical Society. Series A*, 175(4): 831–861, 2012.

A. E. Abbas. General decompositions of multiattribute utility functions with partial utility independence. *Journal of Multicriteria Decision Analysis*, 17:37–59, 2010.

A. E. Abbas and R. A. Howard. Attribute dominance utility. *Decision Analysis*, 2(4): 185–206, 2005.

F. Ackermann. Participants' perceptions on the role of facilitators using group decision support systems. *Group Decision and Negotiation*, 5(1):93–112, 1996.

P. A. Aguilera, A. Fernández, R. Fernández, R. Rumí, and A. Salmerón. Bayesian networks in environmental modelling. *Environmental Modelling & Software*, 26(12): 1376–1388, 2011.

M. Ahlbrecht, R. Borodin, V. Borzenko, J. Ehrhardt, S. French, V. Shershakov, A. Sohier, E. Trakhtengerts, and A. Verbruggen. Decision support issues in RODOS: the needs of decision makers. *Radiation Protection Dosimetry*, 73(1-4):81–84, 1997.

I. Albert, S. Donnet, C. Guihenneuc-Jouyaux, S. Low-Choy, K. Mengersen, and J. Rousseau. Combining expert opinions in prior elicitation. *Bayesian Analysis*, 7 (3):503–532, 2012.

O. Anacleto and C. M. Queen. Dynamic chain graph models for multivariate time series. Tech. Report, Open University, 2013.

O. Anacleto, C. M. Queen, and C. J. Albers. Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors. *Journal of the Royal Statistical Society: Series C*, 62(2):251–270, 2013.

F. J. Anscombe and R. J Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34(1):199–205, 1963.

N. Argyris, A. Morton, and J. R. Figueira. CUT: a multicriteria approach for concavifiable preferences. *Operations Research*, 62(3):633–642, 2014.

K. J. Arrow. *Social choice and individual values.* Yale University Press, 1951.

F. Bacchus and A. Grove. Graphical models for preference and utility. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pages 3–10, 1995.

M. Bacharach. Group decisions in the face of differences of opinion. *Management Science*, 22(2):182–191, 1975.

L. M. Barclay, J. L. Hutton, and J. Q. Smith. Refining a Bayesian network using a chain event graph. *International Journal of Approximate Reasoning*, 54(9):1300–1309, 2013.

L. M. Barclay, J. L. Hutton, and J. Q. Smith. Chain event graphs for informed missingness. *Bayesian Analysis*, 9(1):53–76, 2014.

M. J. Barons, X. Zhong, and J. Q. Smith. Dynamic Bayesian networks for decision support and sugar food security. Technical Report 14-18, CRISM, The University of Warwick, 2014.

D. E. Bell. Multiattribute utility functions: decompositions using interpolation. *Management Science*, 25(8):744–753, 1979.

V. Belton and T. Stewart. *Multiple criteria decision analysis: an integrated approach.* Springer, 2002.

A. Bennet and D. Bennet. The decision-making process in a complex situation. In *Handbook on Decision Support Systems 1*, pages 3–20. 2009.

J. M. Bernardo and A. F. M. Smith. *Bayesian theory.* John Wiley & Sons, 2009.

N. Best, S. Richardson, and A. Thomson. A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14(1):35–59, 2005.

G. Bhandari, K. Hassanein, and R. Deaves. Debiasing investors with decision support systems: an experimental investigation. *Decision Support Systems*, 46(1):399–410, 2008.

D. Bhattacharjya and R. D. Shachter. Evaluating influence diagrams with decision circuits. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, pages 9–16, 2007.

D. Bhattacharjya and R. D. Shachter. Formulating asymmetric decision problems as decision circuits. *Decision Analysis*, 9(2):138–145, 2012.

C. Bielza, M. Gómez, and P. P. Shenoy. A review of representation issues and modeling challenges with influence diagrams. *Omega*, 39(3):227–241, 2011.

K. A. Bollen and J. S. Lang. *Testing structural equation models*. SAGE Publications, 1993.

R. F. Bordley. Combining the opinions of experts who partition events differently. *Decision Analysis*, 6(1):38–46, 2009.

C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-specific independence in Bayesian networks. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence*, pages 115–123, 1996.

X. Boyen and D. Koller. Tractable inference for complex stochastic processes. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 33–42, 1998.

B. Brandherm and A. Jameson. An extension of the differential approach for Bayesian network inference to dynamic Bayesian networks. *International Journal of Intelligent Systems*, 19(8):727–748, 2004.

D. Braziunas and C. Boutilier. Local utility elicitation in GAI models. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, pages 42–49, 2005.

D. R. Brillinger. The calculation of cumulants via conditioning. *Annals of the Institute of Statistical Mathematics*, 21(1):215–218, 1969.

B. G. Buchanan and E. H. Shortliffe. *Rule-based expert systems*. Addison-Wesley Reading, 1984.

W. J. Burns and R. T. Clemen. Covariance structure models and influence diagrams. *Management Science*, 39(7):816–834, 1993.

M. J. Caley, R. A. O'Leary, R. Fisher, S. Low-Choy, S. Johnson, and K. Mengersen. What is an expert? A systems perspective on expertise. *Ecology and Evolution*, 4(3): 231–42, 2014.

G. Caminada, S. French, K. Politis, and J. Q. Smith. Uncertainty in RODOS. Technical Report 94, RODOS, 2000.

Y. B. Canbolat, K. Chelst, and N. Garg. Combining decision tree and MAUT for selecting a country for a global manufacturing facility. *Omega*, 35(3):312–325, 2007.

E. Carter and S. French. Are current processes for nuclear emergency management in Europe adequate? *Journal of Radiological Protection: Official Journal of the Society for Radiological Protection*, 26(4):405–14, 2006.

G. Casella and R. L. Berger. *Statistical inference*. Duxbury, 2002.

E. Castillo. *Expert systems and probabilistic network models*. Springer, 1997.

E. Castillo and U. Kjærulff. Sensitivity analysis in Gaussian Bayesian networks using a symbolic-numerical technique. *Reliability Engineering & System Safety*, 79(2):139–148, 2003.

E. Castillo, J. M. Gutierrez, and A. S. Hadi. Parametric structure of probabilities in Bayesian networks. In *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 89–98. 1995.

E. Castillo, J. M. Gutierrez, and A. S. Hadi. A new method for efficient symbolic propagation in discrete Bayesian networks. *Networks*, 28(1):31–43, 1996.

E. Castillo, J. M. Gutierrez, and A. S. Hadi. Sensitivity analysis in discrete Bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 27(4):412–423, 1997a.

E. Castillo, J. M. Gutierrez, A. S. Hadi, and C. Solares. Symbolic propagation and sensitivity analysis in Gaussian Bayesian networks with application to damage assessment. *Artificial Intelligence in Engineering*, 11(2):173–181, 1997b.

H. Chan and A. Darwiche. When do numbers really matter? In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pages 65–74, 2001.

H. Chan and A. Darwiche. Sensitivity analysis in Bayesian networks: from single to multiple parameters. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, pages 67–75, 2004.

K. M. Chandy, J. Kiniry, A. Rifkin, and D. Zimmerman. A framework for structured distributed object computing. *Parallel Computing*, 24(12):1901–1922, 1998.

R. T. Clemen. *Making hard decisions: an introduction to decision analysis*. Duxbury, 1996.

R. T. Clemen and R. L. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):187–203, 1999.

R. T. Clemen, G. W. Fischer, and R. L. Winkler. Assessing dependence: some experimental results. *Management Science*, 46(8):1100–1115, 2000.

S. Conti, J. P. Gosling, J. E. Oakley, and A. O'Hagan. Gaussian process emulation of dynamic computer codes. *Biometrika*, 96(3):663–676, 2009.

R. M. Cooke. *Experts in uncertainty: opinion and subjective probability in science.* Oxford University Press, 1991.

R. M. Cooke and L. H. J. Goossens. Procedures guide for structural expert judgement in accident consequence modelling. *Radiation Protection Dosimetry*, 90(3):303–309, 2000.

G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2-3):393–405, 1990.

G. F. Cooper and C. Yoo. Causal discovery from a mixture of experimental and observational data. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, 1999.

L. Costa, J. Q. Smith, T. Nichols, J. Cussens, E. P. Duff, and T. R. Makin. Searching multiregression dynamic models of resting-state fMRI networks using integer programming. *Bayesian Analysis (to appear)*, 2015.

R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic networks and expert systems.* Springer-Verlag, 1999.

R. G. Cowell, P. A. Thwaites, and J. Q. Smith. Decision making with decision event graphs. Technical Report 10-15, CRISM, The University of Warwick, 2010.

R. G. Cowell, S. L. Lauritzen, and J. Mortera. Probabilistic expert systems for handling artifacts in complex DNA mixtures. *Forensic Science International: Genetics*, 5(3): 202–209, 2011.

D. A. Cox, J. Little, and D. O'Shea. *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra.* Springer, 2007.

F. G. Cozman. Credal networks. *Artificial intelligence*, 120(2):199–233, 2000.

P. S. Craig, M. Goldstein, J. C. Rougier, and A. H. Seheult. Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96(454):717–729, 2001.

N. Dalkey and O. Helmer. An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3):458–467, 1963.

A. Daneshkhah and J. Q. Smith. Hypercausality, randomisation, and local and global independence. In *Advances in Bayesian Networks*, pages 1–18. Springer, 2004.

A. Darwiche. A differential approach to inference in Bayesian networks. *Journal of the ACM*, 50(3):280–305, 2003.

A. P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B*, 41(1):1–31, 1979.

A. P. Dawid. Separoids: a mathematical framework for conditional independence and irrelevance. *Annals of Mathematical Statistics*, 32:335–372, 2001.

A. P. Dawid. Influence diagrams for causal modelling and inference. *International Statistical Review*, 70(2):161–189, 2002.

A. P. Dawid and P. Constantinou. A formal treatment of sequential ignorability. *Statistics in Biosciences*, 6(2):166–188, 2014.

A. P. Dawid and V. Didelez. Identifying the consequences of dynamic treatment strategies: a decision-theoretic overview. *Statistics Surveys*, 4:184–231, 2010.

A. P. Dawid and S. L. Lauritzen. Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 21(3):1272–1317, 1993.

S. De and Á. E. Faria. Dynamic spatial Bayesian models for radioactivity deposition. *Journal of Time Series Analysis*, 32(6):607–617, 2011.

M. H. de Groot. *Optimal statistical decisions*. John Wiley & Sons, 1970.

R. Demirer and P. P. Shenoy. Sequential valuation networks for asymmetric decision problems. *European Journal of Operational Research*, 169(1):286–309, 2006.

A. P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society. Series B*, 30(2):205–247, 1968.

P. Diaconis and B. Sturmfels. Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, 26(1):363–397, 1998.

P. Diggle and P. J. Ribeiro. *Model-based geostatistics*. Springer Science & Business Media, 2007.

L. Dodd and J. Q. Smith. Devolving command decisions in complex operations. *Journal of the Operational Research Society*, 64(1):17–33, 2012.

M. Drton. Discrete chain graph models. *Bernoulli*, 15(3):736–753, 2009.

M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on algebraic statistics*. Springer, 2009.

J. Ehrhardt, J. Pasler-Sauer, O. Schule, G. Benz, M. Rafat, and J. Richter. Development of RODOS, a comprehensive decision support system for nuclear emergencies in Europe - an overview. *Radiation Protection Dosimetry*, 50(2-4):195–203, 1993.

M. Eichler and V. Didelez. Causal reasoning in graphical time series models. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, pages 109–116, 2007.

Y. Engel and M. P. Wellman. CUI networks: a graphical representation for conditional utility independence. *Journal of Artificial Intelligence Research*, 31:83–112, 2008.

Á. E. Faria and J. Q. Smith. Conditional external Bayesianity in decomposable influence diagrams. In *Bayesian Statistics 5*, pages 551–560. 1996.

Á. E. Faria and J. Q. Smith. Conditionally externally Bayesian pooling operators in chain graphs. *The Annals of Statistics*, 25(4):1740–1761, 1997.

P. H. Farquhar. A fractional hypercube decomposition theorem for multiattribute utility functions. *Operations Research*, 23(5):941–967, 1975.

A. C. Farr, D. Simpson, F. Ruggeri, and K. Mengersen. Combining opinions for use in Bayesian networks: a measurement error approach. Technical report, ePrints, Queensland University of Technology, 2015.

P. C. Fishburn. Interdependence and additivity in multivariate, unidimensional expected utility theory. *International Economic Review*, 8(3):335–342, 1967.

G. Freeman and J. Q. Smith. Bayesian MAP model selection of chain event graphs. *Journal of Multivariate Analysis*, 102(2):1152–1165, 2011.

S. French. Updating of belief in the light of someone else's opinion. *Journal of the Royal Statistical Society. Series A*, 143(1):43–48, 1980.

S. French. Uncertainty and imprecision: modelling and analysis. *Journal of the Operational Research Society*, 46(1):70–79, 1995.

S. French. Uncertainty modelling, data assimilation and decision support for management of off-site nuclear emergencies. *Radiation Protection Dosimetry*, (1):11–15, 1997.

S. French. Aggregating expert judgement. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 105(1):181–206, 2011.

S. French and D. Rios Insua. *Statistical decision theory*. Wiley, 2000.

S. French and B. S. West. Modelling, making inferences and making decisions: the roles of sensitivity analysis. *Top*, 11(2):229–251, 2003.

S. French, R. M. Cooke, and F. Vogt. The use of expert judgement in the context of a physical model. In *Bayesian Statistics 4*, pages 617–624. 1991.

S. French, J. Maule, and K. N. Papamichail. *Decision behaviour, analysis and support.* Cambridge University Press, 2009.

M. Frydenberg. The chain graph Markov property. *Scandinavian Journal of Statistics*, 17:333–353, 1990.

D. M. Gabbay. *Theoretical foundations for non-monotonic reasoning in expert systems.* Springer, 1985.

L. D. Garcia, M. Stillman, and B. Sturmfels. Algebraic geometry of Bayesian networks. *Journal of Symbolic Computation*, 39(3-4):331–355, 2005.

D. Geiger and D. Heckerman. A characterization of the Dirichlet distribution through global and local parameter independence. *The Annals of Statistics*, 25(3):1344–1369, 1997.

D. Geiger, C. Meek, and B. Sturmfels. On the toric algebra of graphical models. *The Annals of Statistics*, 34(3):1463–1492, 2006.

J. Geldermann, V. Bertsch, M. Treitz, S. French, K. N. Papamichail, and R. Hamalainen. Multicriteria decision support and evaluation of strategies for nuclear remediation management. *Omega*, 37(1):238–251, 2009.

A. Gelman and J. Hill. *Data analysis using regression and multilevel/hierarchical models.* Cambridge University Press, 2006.

C. Genest. A characterization theorem for externally Bayesian groups. *The Annals of Statistics*, 12(3):1100–1105, 1984.

C. Genest, K. J. McConway, and M. J. Schervish. Characterization of externally Bayesian pooling operators. *The Annals of Statistics*, 14(2):487–501, 1986.

F. Gering, B. Gerich, E. Wirth, and G. Kirchner. Potential consequences of the Fukushima accident for off-site nuclear emergency management: a case study for Germany. *Radiation Protection Dosimetry*, 155(2):146–154, 2013.

P. Gibilisco, E. Riccomagno, M. P. Rogantin, and H. P. Wynn. *Algebraic and geometric methods in statistics.* Cambridge University Press, 2010.

W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov chain Monte Carlo in practice.* Chapman & Hall, 1996.

M. Goldstein and A. O'Hagan. Bayes linear sufficiency and systems of expert posterior assessments. *Journal of the Royal Statistical Society. Series B*, 58(2):301–316, 1996.

M. Goldstein and J. Rougier. Bayes linear calibrated prediction for complex systems. *Journal of the American Statistical Association*, 101(475):1132–1143, 2006.

M. Goldstein and D. Wooff. *Bayes linear statistics, theory & methods*. Wiley, 2007.

M. Gómez, C. Bielza, J. A. Fernández del Pozo, and S. Ríos-Insua. A graphical decision-theoretic model for neonatal jaundice. *Medical Decision Making*, 27(3):250–265, 2007.

C. Görgen, M. Leonelli, and J. Q. Smith. A differential approach for staged trees, 2015. Accepted to ECSQARU2015.

S. Greenland, J. M. Robins, and J. Pearl. Confounding and collapsibility in causal inference. *Statistical Science*, 14(1):29–46, 1999.

P. J. Harrison and C. F. Stevens. Bayesian forecasting. *Journal of the Royal Statistical Society. Series B*, 38(3):205–247, 1976.

R. G. Haylock and A. O'Hagan. On inference for outputs of computationally expensive algorithms with uncertainty on the inputs. In *Bayesian Statistics 5*, pages 629–637. 1996.

D. Heckerman. Probabilistic similarity networks. *Networks*, 20(5):607–636, 1990.

D. Heckerman, A. Mamdani, and M. P. Wellman. Real-world applications of Bayesian networks. *Communications of the ACM*, 38(3):24–26, 1995.

J. P. T. Higgins and S. Green. *Cochrane handbook for systematic reviews of interventions*. Wiley Online Library, 2008.

S. Hoe and H. Müller. ARGOS - a decision support system for nuclear emergencies. IAEA - INIS, 2003.

R. A. Howard and J. E. Matheson. Influence diagrams. In *Readings on the principles and applications of decision analysis*, pages 719–763. 1983.

R. A. Howard and J. E. Matheson. Influence diagrams. *Decision Analysis*, 2(3):127–143, 2005a.

R. A. Howard and J. E. Matheson. Influence diagram retrospective. *Decision Analysis*, 2(3):144–147, 2005b.

M. Jaeger. Probabilistic decision graphs - combining verification and AI techniques for probabilistic inference. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12:19–42, 2004.

F. Jensen, F. V. Jensen, and S. L. Dittmer. From influence diagrams to junction trees. In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, pages 367–373, 1994.

F. V. Jensen and T. D. Nielsen. *Bayesian networks and decision graphs.* Springer, 2009.

F. V. Jensen and T. D. Nielsen. Probabilistic decision graphs for optimization under uncertainty. *Annals of Operations Research*, 204(1):223–248, 2013.

F. V. Jensen and M. Vomlelová. Unconstrained influence diagrams. In *Proceedings of the 18th Conference on Artificial Intelligence*, pages 234–241, 2002.

F. V. Jensen, T. D. Nielsen, and P. P. Shenoy. Sequential influence diagrams: a unified asymmetry framework. *International Journal of Approximate Reasoning*, 42(1):101–118, 2006.

C. P. Jewell, T. Kypraios, P. Neal, and G. O. Roberts. Bayesian analysis for emerging infectious diseases. *Bayesian Analysis*, 4(3):465–496, 2009.

R. Jirousek and J. Vejnarova. Construction of multidimensional models by operators of composition: current state of art. *Soft Computing*, 7(5):328–335, 2003.

S. Johnson and K. Mengersen. Integrated Bayesian network framework for modeling complex ecological issues. *Integrated Environmental Assessment and Management*, 8 (3):480–90, 2012.

S. Johnson, S. Low-Choy, and K. Mengersen. Integrating Bayesian networks and geographic information systems: good practice examples. *Integrated Environmental Assessment and Management*, 8(3):473–479, 2012.

S. Johnson, L. Marker, K. Mengersen, C. H. Gordon, J. Melzheimer, A. Schmidt-Juntzel, M. Nghikembua, E. Fabiano, J. Henghali, and B. Wachter. Modeling the viability of the free-ranging cheetah population in Namibia: an object-oriented Bayesian network approach. *Ecosphere*, 4(7), 2013.

S. Johnson, E. Abal, K. Ahern, and G. Hamilton. From science to management: using Bayesian networks to learn about Lyngbya. *Statistical Science*, 29(1):36–41, 2014.

M. I. Jordan. Graphical models. *Statistical Science*, 19(1):140–155, 2004.

D. Kahneman and A. Tversky. Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

R. L. Keeney. A group preference axiomatization with cardinal utility. *Management Science*, 23(2):140–145, 1976.

R. L. Keeney. *Value-focused thinking: a path to creative decision making.* Harvard University Press, 1992.

R. L. Keeney. Foundations for group decision analysis. *Decision Analysis*, 10(2):103–120, 2013.

R. L. Keeney and H. Raiffa. *Decision with multiple objectives: preferences and value trade-offs.* Cambridge University Press, 1976.

M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.

M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B*, 63(3):425–464, 2001.

M. C. Kennedy, C. W. Anderson, S. Conti, and A. O'Hagan. Case studies in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety*, 91 (10-11):1301–1309, 2006.

G. Klir and B. Yuan. *Fuzzy sets and fuzzy logic.* Prentice Hall, 1995.

D. Koller and U. Lerner. Sampling in factored dynamic systems. In *Sequential Monte Carlo Methods in Practice*, pages 445–464. Springer-Verlag, 2001.

D. Koller and A. Pfeffer. Object-oriented Bayesian networks. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, pages 302–313, 1997.

K. B. Korb and A. E. Nicholson. *Bayesian artificial intelligence.* CRC Press, 2003.

T. N. Krishnamurti, C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran. Multimodel ensemble forecasts for weather and seasonal climate. *Journal of Climate*, 13(23):4196–4216, 2000.

P. La Mura and Y. Shoham. Expected utility networks. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, pages 366–373, 1999.

K. B. Laskey. MEBN: a language for first-order Bayesian knowledge bases. *Artificial Intelligence*, 172(2):140–178, 2008.

K. B. Laskey and P. C. G. da Costa. Uncertainty representation and reasoning in complex systems. In *Complex Systems in Knowledge-based Environments: Theory, Models and Applications*, pages 7–39. 2009.

K. B. Laskey and S. M. Mahoney. Network fragments: representing knowledge for constructing probabilistic models. In *Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence*, pages 334–341, 1997.

K. B. Laskey, E. J. Wright, and P. C. G. da Costa. Envisioning uncertainty in geospatial information. *International Journal of Approximate Reasoning*, 51(2):209–223, 2010.

S. L. Lauritzen. Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association*, 87(420):1098–1108, 1992.

S. L. Lauritzen. *Graphical Models.* Oxford University Press, 1996.

S. L. Lauritzen. Causal inference from graphical models. *Complex stochastic systems*, pages 63–107, 2001.

S. L. Lauritzen and T. S. Richardson. Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society: Series B*, 64(3):321–348, 2002.

M. Leonelli and J. Q. Smith. Using graphical models and multi-attribute utility theory for probabilistic uncertainty handling in large systems, with application to the nuclear emergency management. In *29th International Conference on Data Engineering Workshops (ICDEW)*, pages 181–192, 2013a.

M. Leonelli and J. Q. Smith. Dynamic uncertainty handling for coherent decision making in nuclear emergency response. In *Risk Management for Complex Socio-Technical Systems*, 2013b. To appear.

M. Leonelli and J. Q. Smith. Bayesian decision support for complex systems with many distributed experts, 2015. Invited revision to the Annals of Operations Research.

M. Leonelli, E. Riccomagno, and J. Q. Smith. Using computer algebra to symbolically evaluate discrete influence diagrams, 2015a. Submitted.

M. Leonelli, E. Riccomagno, and J. Q. Smith. The algebra of integrated partial belief systems, 2015b. In preparation.

D. V. Lindley. Reconciliation of discrete probability distributions. In *Bayesian Statistics 2*, pages 375–390. 1985.

F. Lindsten, A. M. Johansen, C. A. Naesseth, B. Kirkpatrick, T. B. Schön, J. Aston, and A. Bouchard-Côté. Divide-and-conquer with sequential monte carlo. *arXiv:1406.4993*, 2014.

H. A. Linstone and M. Turoff. *The Delphi method: techniques and applications*. Addison-Wesley Reading, 1975.

F. Liu and M. West. A dynamic modelling strategy for Bayesian computer model emulation. *Bayesian Analysis*, 4(2):393–411, 2009.

A. Madansky. Externally Bayesian groups. Technical report, The RAND Corporation, 1964.

A. L. Madsen and F. V. Jensen. Lazy evaluation of symmetric Bayesian decision problems. In *Proceedings of the 15th Conference of Uncertainty in Artificial Intelligence*, pages 382–390, 1999.

A. L. Madsen and F. V. Jensen. Solving linear-quadratic conditional Gaussian influence diagrams. *International Journal of Approximate Reasoning*, 38(3):263–282, 2005.

S. M. Mahoney and K. B. Laskey. Network engineering for complex belief networks. In *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence*, pages 389–396, 1996.

G. M. Marakas. *Decision support systems in the 21st century.* Prentice Hall, 2003.

M. S. Massa and S. L. Lauritzen. Combining statistical models. *Contemporary Mathematics*, 516:239–259, 2010.

P. McCullagh. *Tensor methods in statistics.* Chapman and Hall, 1987.

T. J McKinley, J. V. Ross, R. Deardon, and A. R. Cook. Simulation-based Bayesian inference for epidemic models. *Computational Statistics & Data Analysis*, 71:434–447, 2014.

J. R. W. Merrick. Getting the right mix of experts. *Decision Analysis*, 5(1):43–52, 2008.

T. Mikkelsen, S. Galmarini, R. Bianconi, and S. French. ENSEMBLE: methods to reconcile disparate national long range dispersion forecasts. Technical report, RISO, Roskilde, 2003.

L. Monte, J. E. Brittain, E. Gallego, L. Håkanson, D. Hofman, and A. Jiménez. MOIRA-PLUS: a decision support system for the management of complex fresh water ecosystems contaminated by radionuclides and heavy metals. *Computers & Geosciences*, 35 (5):880–896, 2009.

J. Mortera, P. Vicard, and C. Vergara. Object-oriented Bayesian networks for a decision support system for antitrust. *The Annals of Applied Statistics*, 7(2):714–738, 2013.

S. M. Müller and M. J. Machina. Moment preferences and polynomial utility. *Economics Letters*, 23(4):349–353, 1987.

J. L. Mumpower and T. R. Stewart. Expert judgement and expert disagreement. *Thinking & Reasoning*, 2(2-3):191–212, 1996.

K. P. Murphy. *Dynamic Bayesian networks: representation, inference and learning.* PhD thesis, University of California, Berkeley, 2002.

K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: an empirical study. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, pages 467–475, 1999.

S. A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B*, 65(2):331–355, 2003.

R. E. Neapolitan. *Learning Bayesian networks.* Prentice Hall, 2004.

M. Neil, N. Fenton, and L. Nielson. Building large-scale Bayesian networks. *The Knowledge Engineering Review*, 15(03):257–284, 2000.

D. Niedermayer. An introduction to Bayesian networks and their contemporary applications. In *Innovations in Bayesian Networks*, pages 117–130. Springer, 2008.

T. D. Nielsen and F. V. Jensen. Representing and solving asymmetric decision problems. *International Journal of Information Technology & Decision Making*, 2(02):217–263, 2003.

D. Nilsson. The computation of moments of decomposable functions in probabilistic expert systems. In *Proceedings of the 3rd International Symposium on Adaptive Systems*, pages 116–121, 2001.

C. J Oates. *Bayesian inference for protein signalling networks.* PhD thesis, The University of Warwick, 2013.

A. O'Hagan. Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering & System Safety*, 91(10-11):1290–1300, 2006.

A. O'Hagan and J. Forster. *Bayesian inference.* Arnold, 2004.

A. O'Hagan, C.E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. *Uncertain judgements: eliciting experts' probabilities.* John Wiley & Sons, 2006.

R. M. Oliver and J. Q. Smith. *Influence diagrams, belief nets, and decision analysis.* John Wiley & Sons, 1990.

S. M. Olmsted. *On representing and solving decision problems.* PhD thesis, Stanford University, 1984.

L. Pachter and B. Sturmfels. *Algebraic statistics for computational biology.* Cambridge University Press, 2005.

K. N. Papamichail and S. French. Explaining and justifying the advice of a decision support system: a natural language generation approach. *Expert Systems with Applications*, 24(1):35–48, 2003.

K. N. Papamichail and S. French. Design and evaluation of an intelligent decision support system for nuclear emergencies. *Decision Support Systems*, 41(1):84–111, 2005.

K. N. Papamichail and S. French. 25 Years of MCDA in nuclear emergency management. *IMA Journal of Management Mathematics*, 24(4):481–503, 2013.

P. A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization.* PhD thesis, California Institute of Technology, 2000.

J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann, 1988.

J. Pearl. *Causality: models, reasoning and inference.* Cambridge University Press, 2000.

D. M. Pennock and M. P. Wellman. Graphical models for groups: belief aggregation and risk sharing. *Decision Analysis*, 2(3):148–164, 2005.

D. Peteiro-Barral and B. Guijarro-Berdiñas. A survey of methods for distributed machine learning. *Progress in Artificial Intelligence*, 2(1):1–11, 2013.

L. D. Phillips. A theory of requisite decision models. *Acta Psychologica*, 56(1):29–48, 1984.

L. D. Phillips and M. C. Phillips. Facilitated work groups: theory and practice. *Journal of the Operational Research Society*, 44:533–549, 1993.

G. Pistone and H. P. Wynn. Generalised confounding with Gröbner bases. *Biometrika*, 83(3):653–666, 1996.

G. Pistone, E. Riccomagno, and H. P. Wynn. Gröbner bases and factorisation in discrete probability and Bayes. *Statistics and Computing*, 11:37–46, 2001.

G. Pistone, E. Riccomagno, and H. P. Wynn. *Algebraic statistics: computational commutative algebra in statistics.* CRC Press, 2002.

T. Poston and I. Stewart. *Catastrophe theory and its applications.* Courier Corporation, 2014.

C. M. Queen. Using the multiregression dynamic model to forecast brand sales in a competitive product market. *The Statistician*, 43(1):87–98, 1994.

C. M. Queen and C. J. Albers. Intervention and causality: forecasting traffic flows using a dynamic Bayesian network. *Journal of the American Statistical Association*, 104 (486):669–681, 2009.

C. M. Queen and J. Q. Smith. Dynamic graphical models. In *Bayesian Statistics 4*, pages 741–751. 1992.

C. M. Queen and J. Q. Smith. Multiregression dynamic models. *Journal of the Royal Statistical Society. Series B*, 55(4):849–870, 1993.

C. M. Queen, B. J. Wright, and C. J. Albers. Forecast covariances in the linear multi-regression dynamic model. *Journal of Forecasting*, 27(2):175–191, 2008.

W. Raskob, F. Gering, and V. Bertsch. Approaches to visualisation of uncertainties to decision makers in an operational decision support system. In *Proceedings of the 6th International ISCRAM Conference*, 2009.

S. Renooij. Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review*, 16(03):255–269, 2001.

E. Riccomagno. A short history of algebraic statistics. *Metrika*, 69(2-3):397–418, 2009.

T. Richardson and P. Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30 (4):962–1030, 2002.

M. Rodríguez, D. M. Escalante, and A. Peregrín. Efficient distributed genetic algorithm for rule extraction. *Applied Soft Computing*, 11(1):733–743, 2011.

J. Rougier. Efficient emulators for multivariate deterministic functions. *Journal of Computational and Graphical Statistics*, 17(4):827–843, 2008.

J. Rougier and M. Crucifix. Uncertainty in climate science and climate policy. *arXiv:1411.6878*, 2014.

T. J. Santner, B. J. Williams, and W. Notz. *The design and analysis of computer experiments.* Springer, 2003.

L. J. Savage. *The foundations of statistics.* Courier Dover Publications, 1972.

R. Settimi and J. Q. Smith. Geometry, moments and conditional independence trees with hidden variables. *Annals of Statistics*, 28(4):1179–1205, 2000.

R. D. Shachter. Evaluating influence diagrams. *Operations Research*, (6):871–882, 1986.

R. D. Shachter and C. R. Kenley. Gaussian influence diagrams. *Management Science*, 35(5):527–550, 1989.

G. R. Shafer. *A mathematical theory of evidence.* Princeton University Press, 1976.

G. R. Shafer. *The art of causal conjecture.* MIT Press, 1996.

V. Siskind. Second moments of inverse Wishart matrix elements. *Biometrika*, 59(3): 690–691, 1972.

J. E. Smith, S. Holtzman, and J. E. Matheson. Structuring conditional relationships in influence diagrams. *Operations Research*, 41(2):280–297, 1993.

J. Q. Smith. Influence diagrams for Bayesian decision analysis. *European Journal of Operational Research*, 40(3):363–376, 1989a.

J. Q. Smith. Influence diagrams for statistical modelling. *The Annals of Statistics*, 17 (2):654 – 672, 1989b.

J. Q. Smith. Plausible Bayesian games. In *Bayesian Statistics 5*, pages 387–406. 1996.

J. Q. Smith. *Bayesian decision analysis: principles and practice.* Cambridge University Press, 2010.

J. Q. Smith and P. E. Anderson. Conditional independence and chain event graphs. *Artificial Intelligence*, 172(1):42–68, 2008.

J. Q. Smith and L. Dodd. Regulating autonomous agents facing conflicting objectives: a command and control example. *Decision Analysis*, 9(2):165–171, 2012.

J. Q. Smith and Á. E. Faria. Bayesian Poisson models for the graphical combination of dependent expert information. *Journal of the Royal Statistical Society. Series B*, 62 (3):525–544, 2000.

J. Q. Smith and L. J. Figueroa. A causal algebra for dynamic flow networks. In *Advances in Probabilistic Graphical Models*, pages 39–54. Springer, 2007.

J. Q. Smith and P. Thwaites. Influence diagrams. In *Encyclopaedia of Quantitative Risk Analysis and Assessment*, pages 897–910. 2008.

J. Q. Smith, Á. E. Faria, S. French, D. Ranyard, D. Vlesshhouwer, J. Bohunova, T. Duranova, M. Stubna, L. Dutton, C. Rojas, and A. Sohier. Probabilistic data assimilation within RODOS. *Radiation Protection Dosimetry*, 73(1):57–59, 1997.

J. Q. Smith, M. Barons, and M. Leonelli. Coherent inference for integrating decision support systems, 2015. In preparation.

D. J. Spiegelhalter and S. L. Lauritzen. Sequential updating of conditional probabilities on directed graphical structures. *Networks*, 20(5):579–605, 1990.

D. J. Spiegelhalter, A. P. Dawid, S. L. Lauritzen, and R. G. Cowell. Bayesian analysis in expert systems. *Statistical Science*, 8(3):219–247, 1993.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, prediction, and search.* Lecture Notes in Statistics. Springer, 1993.

A. Stassopoulou, M. Petrou, and J. Kittler. Application of a Bayesian network in a GIS based decision making system. *International Journal of Geographical Information Science*, 12(1):23–46, 1998.

M. Stone. The opinion pool. *The Annals of Mathematical Statistics*, 32(4):1339–1342, 1961.

S. Sullivant. Algebraic Geometry of Gaussian Bayesian Networks. *Advances in Applied Mathematics*, 40(4):482–513, 2008.

S. Sullivant, K. Talaska, and J. Draisma. Trek separation for Gaussian graphical models. *The Annals of Statistics*, 38(3):1665–1685, 2010.

P. Thwaites. Causal identifiability via chain event graphs. *Artificial Intelligence*, 195: 291–315, 2013.

P Thwaites, J. Q. Smith, and R. G. Cowell. Propagation using chain event graphs. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, pages 546–553, 2008.

P. A. Thwaites and J. Q. Smith. A separation theorem for chain event graphs. *arXiv:1501.05215*, 2015.

P. A. Thwaites, J. Q. Smith, and E. Riccomagno. Causal analysis with chain event graphs. *Artificial Intelligence*, 174(12-13):889–909, 2010.

D. B. Turner. *Workbook of atmospheric dispersion estimates: an introduction to dispersion modeling.* CRC press, 1994.

A. Tversky and D. Kahneman. Judgment under uncertainty: heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

J. B. Ullman and P. M. Bentler. *Structural equation modeling.* Wiley Online Library, 2003.

L. Uusitalo. Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, 203(3):312–318, 2007.

C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. *Statistica Sinica*, 21(1):5–42, 2011.

J. Von Neumann and O. Morgenstern. *Theory of games and economic behaviour*. Princeton University Press, 1947.

D. Von Winterfeldt and W. Edwards. *Decision analysis and behavioral research*. Cambridge University Press, 1986.

P. P. Wakker. Explaining the characteristics of the power (CRRA) utility family. *Health Economics*, 17:1329–1344, 2008.

M. M. Wall and Y. Amemiya. Estimation for polynomial structural equation models. *Journal of the American Statistical Association*, 95(451):929–940, 2000.

B. Walle and M. Turoff. Decision support for emergency situations. *Information Systems and e-Business Management*, 6(3):295–316, 2008.

P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman & Hall, 1991.

N. Wermuth. Analogies between multiplicative models in contingency tables and covariance selection. *Biometrics*, 32(1):95–108, 1976.

M. West and P. J. Harrison. *Bayesian forecasting and dynamic models*. Springer, 1997.

C. K. Wikle, R. F. Milliff, D. Nychka, and L. M. Berliner. Spatiotemporal hierarchical Bayesian modeling tropical ocean surface winds. *Journal of the American Statistical Association*, 96(454):382–397, 2001.

D. Williamson and M. Goldstein. Bayesian policy support for adaptive strategies using computer models for complex physical systems. *Journal of the Operational Research Society*, 63(8):1021–1033, 2011.

R. L. Winkler and R. T. Clemen. Multiple experts vs. multiple methods: combining correlation assessments. *Decision Analysis*, 1(3):167–176, 2004.

M. P. Wiper and S. French. Combining experts' opinions using a normal-Wishart model. *Journal of Forecasting*, 14(1):25–34, 1995.

Y. Xiang. *Probabilistic reasoning in multiagent systems*. Cambridge University Press, 2002.

Y. Xiang, J. Q. Smith, and J. Kroes. Multiagent Bayesian forecasting of structural time-invariant dynamic systems with graphical models. *International Journal of Approximate Reasoning*, 52(7):960–977, 2011.

E. C. Zeeman. *Catastrophe theory*. Springer, 1979.

Z. Y. Zhao and M. West. Dynamic dependence networks: multiregression dynamic models for financial time series analysis and portfolio decisions. Technical report, Department of Statistical Science, Duke University, 2015.

P. Zwiernik and J. Q. Smith. Tree cumulants and the geometry of binary tree models. *Bernoulli*, 18(1):290–321, 2012.