Multi-Context Models for Reasoning under Partial Knowledge: Generative Process and Inference Grammar

Ardavan S. Nobandegani
Dept. of Electrical and Computer Engineering
McGill University
Montreal, QC H3A 0E9

Ioannis N. Psaromiligkos
Dept. of Electrical and Computer Engineering
McGill University
Montreal, QC H3A 0E9

Abstract

Arriving at the complete probabilistic knowledge of a domain, i.e., learning how all variables interact, is indeed a demanding task. In reality, settings often arise for which an individual merely possesses partial knowledge of the domain, and yet, is expected to give adequate answers to a variety of posed queries. That is, although precise answers to some queries, in principle, cannot be achieved, a range of plausible answers is attainable for each query given the available partial knowledge. In this paper, we propose the Multi-Context Model (MCM), a new graphical model to represent the state of partial knowledge as to a domain. MCM is a middle ground between Probabilistic Logic, Bayesian Logic, and Probabilistic Graphical Models. For this model we discuss: (i) the dynamics of constructing a contradiction-free MCM, i.e., to form partial beliefs regarding a domain in a gradual and probabilistically consistent way, and (ii) how to perform inference, i.e., to evaluate a probability of interest involving some variables of the domain.

1 INTRODUCTION

At an abstract level, an individual (also referred to as a reasoner) is faced with a domain where by "domain" we simply mean a collection of propositions or concepts which are mathematically encoded as Random Variables (RVs). To arrive at the complete probabilistic knowledge of the domain, i.e., to learn how all RVs in the domain probabilistically interact with one another, is indeed a demanding task. In reality, an individual is often faced with a domain for which she merely possesses *partial* knowledge—that is, she only knows how *some* (not all) RVs in the domain interact. To make the setting under study more tangible, consider the following case. Suppose that the probabilistic knowledge of a domain is represented by a Probabilis-

tic Graphical Model (PGM) $\mathcal{B}$, e.g., a Bayesian Network (BN). Then the reasoner comes across a new RV, say ψ, and would like to incorporate it into $\mathcal{B}$ so as to achieve the complete probabilistic knowledge of the new domain (which now also includes ψ). However, incorporation of ψ into $\mathcal{B}$ would require knowledge of how ψ is probabilistically related to all the RVs already present in $\mathcal{B}$; a knowledge which may be, quite plausibly, unavailable to the reasoner. An interesting question that now arises is how to handle situations where only partial knowledge as to how ψ is probabilistically related to $\mathcal{B}$ is available. An example would be when the reasoner merely knows how ψ interacts probabilistically with only one RV, say ϕ, in $\mathcal{B}$.

In this paper, a graphical model, namely, the Multi-Context Model (MCM) is proposed to represent the setting in which only partial probabilistic knowledge of a domain is available to the reasoner. More specifically, MCM is a graphical language to represent settings in which the Joint Probability Distribution (JPD) over all RVs is not available, but what is available instead is the JPDs over a collection of subsets of RVs of the domain (referred to as sub-domains or *contexts*). These contexts are potentially overlapping, i.e., they could share some RVs. As pointed out elegantly in (Pearl 1990), *"this state of partial knowledge is more common, because we often begin thinking about a problem through isolated frames, paying no attention to interdependencies."* Along the same line of thought, it is plausible to assume that the probabilistic knowledge of the domain at the early primitive stage consists of a collection of disjoint contexts and as the reasoner acquires more knowledge as to how the variables in the model are related to one another and thus probabilistically interact, contexts gradually go through a process very much like an evolution: contexts start to share some variables, overlaps begin to emerge and, once enough knowledge is obtained, a number of contexts could merge thereby giving rise to bigger contexts. This naturally raises the following fundamental question: How could a collection of consistent, probabilistically sound, and potentially overlapping contexts emerge *gradually* over the course of time? In an attempt to answer this question we present a generative process of constructing a contradiction-free

MCM. Finally, we would like to note that the special case where the whole domain is modeled as a single context corresponds to the conventional way of modeling the probabilistic knowledge of a domain using a single PGM, e.g., by some BN.

Another yet crucial question which we address in this work—which is another motivation behind the development of the MCM—is how the task of inference (i.e., the evaluation of some probability of interest which is hereafter referred to as *query*) should be carried out in a domain which is modeled according to some MCM. A query does not necessarily belong to any one of the contexts in particular and, in fact, may involve RVs from different contexts.

The paper is structured as follows. After introducing the notation in Sec. 2, we define in Sec. 3 the MCM and drawing on the notion of probabilistic conditioning, a generative process of constructing a contradiction-free MCM is discussed. Then, in Sec. 4 we elaborate on the problem of inference in a multi-context setting, i.e., in a domain whose probabilistic knowledge is encoded as an MCM. In Sec. 5 we discuss the relevant past work and comment on the proposed model. Finally, Sec. 6 concludes the paper.

2 TERMINOLOGY AND NOTATION

In this section we present the mathematical notation and the terminology employed in this paper. Random quantities are denoted by bold-faced letters; their realizations are denoted by the same letter but non-bold. More specifically, RVs are denoted by lower-case bold-faced letters, e.g., $\mathbf{x}$, while random vectors are denoted by upper-case bold letters, e.g., $\mathbf{X}$. *Val*$(\cdot)$ denotes the set of values a random quantity can take, e.g., *Val*$(\mathbf{x})$ is the set of all possible realizations of the RV $\mathbf{x}$. In this paper, we assume that all random quantities are discrete.

The JPD over the RVs $\mathbf{x}_1, \cdots, \mathbf{x}_n$ is denoted by $\mathbb{P}(\mathbf{x}_1, \cdots, \mathbf{x}_n)$; when $\mathbf{x}_1, \cdots, \mathbf{x}_n$ comprise a vector $\mathbf{X}$ then $\mathbb{P}(\mathbf{X}) := \mathbb{P}(\mathbf{x}_1, \cdots, \mathbf{x}_n)$. We will use the notation $\mathbf{x}_{1:n}$ to denote the sequence of n RVs $\mathbf{x}_1, \cdots, \mathbf{x}_n$. To simplify presentation and to prevent our expressions from becoming cumbersome, we incur the following abuse of notation: We denote the probability $\mathbb{P}(\mathbf{x} = x)$ by $\mathbb{P}(x)$ for some RV $\mathbf{x}$ and its realization $x \in$ *Val*$(\mathbf{x})$. Also, $\mathbb{P}(\bar{x}) := \mathbb{P}(\mathbf{x} \neq x) = 1 - \mathbb{P}(x)$ for some $x \in$ *Val*$(\mathbf{x})$, i.e., $\mathbb{P}(\bar{x})$ is the probability that $\mathbf{x}$ takes on any value other than x. For conditional probabilities we will use the notation $\mathbb{P}(x|y)$ instead of $\mathbb{P}(\mathbf{x} = x|\mathbf{y} = y)$. Similar notations will be used for the case of random vectors, i.e., $\mathbb{P}(X) := \mathbb{P}(\mathbf{X} = X)$, $\mathbb{P}(\bar{X}) := \mathbb{P}(\mathbf{X} \neq X) = 1 - \mathbb{P}(\mathbf{X} = X) = 1 - \mathbb{P}(X)$, and $\mathbb{P}(X|Y) := \mathbb{P}(\mathbf{X} = X|\mathbf{Y} = Y)$.

The subscript $\downarrow$ on a probability, e.g., $\mathbb{P}(x|y)_{\downarrow}$, denotes the minimum value the probability can take subject to the constraints induced by the available probabilistic knowl-

edge. Likewise, the subscript $\uparrow$ on a probability denotes the maximum value the probability can take. Finally, the operator $[\cdot]^+$ gives the positive part of its argument, i.e., $[a]^+ := \max\{0, a\}$ for any real-valued a.

3 MULTI-CONTEXT MODEL

As explained earlier, a *domain* is simply the set of all Random Variables (RVs) at hand. A *context* comprises a collection of RVs for which their JPD is precisely known, see Fig. 1(a). In general, two contexts could be disjoint (Fig. 1(b)) or overlapping (Fig. 1(c)).
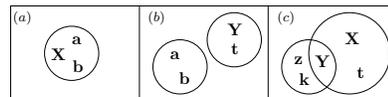

Figure 1: Graphical representation of contexts: (a) Context associated to $\mathbb{P}(\mathbf{a}, \mathbf{b}, \mathbf{X})$. (b) Two disjoint contexts associated to $\mathbb{P}(\mathbf{a}, \mathbf{b})$ and $\mathbb{P}(\mathbf{Y}, \mathbf{t})$. (c) Two overlapping contexts associated to $\mathbb{P}(\mathbf{X}, \mathbf{Y}, \mathbf{t})$ and $\mathbb{P}(\mathbf{Y}, \mathbf{z}, \mathbf{k})$. The random vector $\mathbf{Y}$ is referred to as the *induced* part in Sec. 3.

A *Multi-Context Model (MCM)* encodes the probabilistic knowledge of a domain as a collection of possibly overlapping contexts. This enables the handling of situations in which comprehensive knowledge of a domain is not available, but partial information is, in the form of JPDs of some subsets of the domain. Let us first motivate the proposed MCM by entertaining a simple yet enlightening example.

3.1 MOTIVATING EXAMPLE

Consider a domain consisting of the RVs $\mathbf{y}, \mathbf{z}$ in addition to a set of n RVs, $\mathbf{x}_{1:n}$. A reasoner has formed a partial belief as to the probabilistic connections between the variables of the domain. More specifically, the reasoner knows precisely the JPDs $\mathbb{P}(\mathbf{y}, \mathbf{z})$ and $\mathbb{P}(\mathbf{x}_{1:n})$ but not the JPD $\mathbb{P}(\mathbf{y}, \mathbf{z}, \mathbf{x}_{1:n})$. This setting is described by an MCM that consists of two disjoint contexts, one associated to RVs $\mathbf{y}, \mathbf{z}$ and the other to $\mathbf{x}_{1:n}$, as shown in Fig. 2.
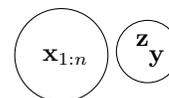

Figure 2: Problem statement as an MCM.

Assume that the following query is posed: Given the available information, what could be said about $\mathbb{P}(y|x_i)$ for some $i = 1, \cdots, n$? The RVs $\mathbf{y}$ and $\mathbf{x}_i$ belong to different contexts, therefore, the JPD of $\mathbf{y}$ and $\mathbf{x}_i$, $\mathbb{P}(\mathbf{x}_i, \mathbf{y})$, is not available. The best one can hope for is to derive the range within which $\mathbb{P}(y|x_i)$ varies, namely, $[\mathbb{P}(y|x_i)_{\downarrow}, \mathbb{P}(y|x_i)_{\uparrow}]$.

Let us for the moment assume the objective is to find $\mathbb{P}(y|x_i)_\downarrow$. Based on the conventional methodology, i.e., the approach adopted by past work (cf. (Andersen and Hooker 1990; 1994; Hansen et al. 1995) and references therein) one has to write down *all* the information as a list of linear equations and solve it as a Linear Program (LP). The main drawback of the conventional approach is that it cannot distinguish between what information is relevant and what is irrelevant for the posed query, and hence what needs to and what need not be considered in answering the query. The price for this is that the number of parameters required to merely formulate the query as an LP is exponential in n.

The key point, however, is that what information is relevant (or irrelevant) depends directly on the posed query, i.e., it is query-dependent. The main advantage of the proposed MCM over previous approaches is that it enables answering a query in a computationally efficient manner by distinguishing the relevant information from the irrelevant for the given query. This is realized thorough adopting the notion of *inference grammar*; a concept which will be systematically defined later. For our example, following the inference rule we will provide in Sec. 4.2, one can easily get $\mathbb{P}(y|x_i)_\downarrow = \left[\frac{\mathbb{P}(y)-\mathbb{P}(\bar{x}_i)}{\mathbb{P}(x_i)}\right]^+$.

The task of inference in an MCM is carried out on two different levels, which makes the task more computationally efficient:

(i) High-Level Reasoning: at this level, through the use of inference grammar, the relevant quantities are identified (e.g., $\mathbb{P}(y)$ and $\mathbb{P}(\bar{x}_i)$ in the case of our example).

(ii) Low-Level Reasoning: the relevant quantities, identified in (i), can then be computed by employing inference algorithms which take advantage of the potentially rich independence structure governing the contexts. For example, it could very well be the case that for the JPD associated to $\mathbf{x}_{1:n}$ a large number of conditional independence relations hold. In that case, stating the derivation of $\mathbb{P}(\bar{x}_i)$ (i.e., $1 - \mathbb{P}(x_i)$) as an LP would be computationally inefficient[1] but unnecessary. Indeed, the task of finding $\mathbb{P}(\bar{x}_i)$ could be accomplished in a computationally efficient way using one of the many inference methods developed for probabilistic graphical models; a key point that the previous approaches do not take advantage of.

As a final step, in order to derive the lower/upper bound to the posed query, the quantities identified in (i) and subsequently calculated in (ii) are stated and solved as an LP.

The idea behind "high-level reasoning" will be explained and clarified further in Sec. 4.2 and 4.3, while the concept of "low-level reasoning" will be discussed in Sec. 4.1.

3.2 GENERATIVE PROCESS OF CONTRADICTION-FREE MCMS

The objective of the generative process we describe in this section is to provide a way to consistently[2] construct contexts, in a sequential manner, over a set of RVs. The act of constructing a context, i.e., of assigning a JPD to a subset of RVs, corresponds to forming a *subjective*[3] belief over those RVs. In this light, the act of constructing multiple contexts corresponds to *gradually* forming subjective beliefs over a number of subsets of variables in the domain; hence every context symbolizes an established belief over the RVs involved in that context.

We introduce this problem by considering a simple case shown in Fig. 3(a). Suppose there are three RVs, namely,
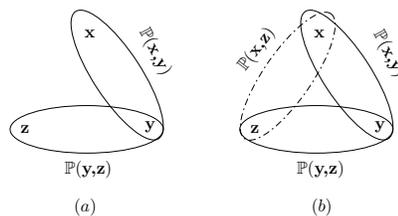

(a) (b)

Figure 3: Generative process for contradiction-free Multi-Context Model. The dash-dotted contexts cannot be freely assigned.

$\mathbf{x}, \mathbf{y}$, and $\mathbf{z}$, present in the domain and let us consider the following question: Could one assign $\mathbb{P}(\mathbf{x}, \mathbf{y})$ and $\mathbb{P}(\mathbf{y}, \mathbf{z})$, *freely* and *gradually* in a consistent manner, over the three variables without introducing any sort of contradiction? It is easy to verify that the answer is positive. Indeed, one could start off by assigning $\mathbb{P}(\mathbf{x}, \mathbf{y})$. This assignment would, of course, induce the marginal $\mathbb{P}(\mathbf{y})$ and one can write $\mathbb{P}(\mathbf{y}, \mathbf{z}) = \mathbb{P}(\mathbf{y})\mathbb{P}(\mathbf{z}|\mathbf{y})$. Then, to complete this task, one would just need to proceed with assigning $\mathbb{P}(\mathbf{z}|\mathbf{y})$. This process could be referred to as a *generative* process of the assignment of $\mathbb{P}(\mathbf{x}, \mathbf{y})$ and $\mathbb{P}(\mathbf{y}, \mathbf{z})$ over $\mathbf{x}, \mathbf{y}$, and $\mathbf{z}$ without introducing any inconsistencies, in a gradual manner. Indeed, free-assignment refers to the act of freely assigning the non-induced, e.g., $P(\mathbf{z}|\mathbf{y})$, part of the *to-be-formed* belief, e.g., $P(\mathbf{y}, \mathbf{z})$. In other words, free-assignment signifies the observation that the already-formed belief does not impose any constraints on the non-induced part of the to-be-formed belief.

---

[1] The number of parameters required just to state the problem as an LP is exponential in n.

[2] That is, without introducing any form of contradictory result with respect to any probability assignment.

[3] One must not interpret the subjectivity of belief as "total disconnectivity from the reality." Thus, we adopt the Bayesian interpretation of probability in this section. The avid reader is referred to (Chalmers 1976). An adherent to the frequentist interpretation of probability could think of contexts as being empirically constructed from a collection of data and thus skip Sec. 3.2 and proceed directly to the next section.

Let us now consider the case shown in Fig. 3(b). Could one assign $\mathbb{P}(\mathbf{x}, \mathbf{y}), \mathbb{P}(\mathbf{y}, \mathbf{z})$, and $\mathbb{P}(\mathbf{x}, \mathbf{z})$ freely and gradually in a consistent manner over the three variables without introducing any sort of contradiction? After some investigation, one can see that the answer is negative (Pearl 1985). Not surprisingly, the reason for this has to do with the existence of a loop in the model: once $\mathbb{P}(\mathbf{x}, \mathbf{y})$ and $\mathbb{P}(\mathbf{y}, \mathbf{z}) = \mathbb{P}(\mathbf{y})\mathbb{P}(\mathbf{z}|\mathbf{y})$ are assigned[4], then $\mathbb{P}(\mathbf{x}, \mathbf{z})$ cannot be assigned freely. This is due to the fact that $\mathbb{P}(\mathbf{x}, \mathbf{z})$ has to satisfy some non-trivial conditions imposed by the already assigned contexts $\mathbb{P}(\mathbf{x}, \mathbf{y})$ and $\mathbb{P}(\mathbf{y}, \mathbf{z})$ (Pearl 1985).

In summary, whenever it comes to generating a new context, the JPD associated to that context has to be separated into two parts: (i) the part induced by the already existing contexts, and (ii) the part containing new variables which have never been so far associated to any context (i.e., non-induced part). The key point in the generation of contradiction-free MCMs is that the former part has to be induced by some context which, itself, is already present in the domain. That is, all the induced parts have to be already contained within some context. Otherwise, to include the induced parts—each constrained by the context it is already in—in a new context, the newly created context would have to satisfy some nontrivial constraints and therefore could not be *freely* assigned.
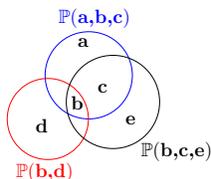

Figure 4: MCM for $\mathbb{P}(\mathbf{a}, \mathbf{b}, \mathbf{c}), \mathbb{P}(\mathbf{b}, \mathbf{d})$, and $\mathbb{P}(\mathbf{b}, \mathbf{c}, \mathbf{e})$.

Let us discuss one final case to further clarify the process. Consider the multi-context model in Fig. 4. Could this model be constructed freely and gradually in a probabilistically consistent manner? The answer is positive. We first assign $\mathbb{P}(\mathbf{a}, \mathbf{b}, \mathbf{c})$, then we assign $\mathbb{P}(\mathbf{b}, \mathbf{c}, \mathbf{e}) = \mathbb{P}(\mathbf{b}, \mathbf{c})\mathbb{P}(\mathbf{e}|\mathbf{b}, \mathbf{c})$ where $\mathbb{P}(\mathbf{b}, \mathbf{c})$ is induced by our first assignment of $\mathbb{P}(\mathbf{a}, \mathbf{b}, \mathbf{c})$. Finally, we assign $\mathbb{P}(\mathbf{b}, \mathbf{d}) = \mathbb{P}(\mathbf{b})\mathbb{P}(\mathbf{d}|\mathbf{b})$ where $\mathbb{P}(\mathbf{b})$ is induced by our first assignment of $\mathbb{P}(\mathbf{a}, \mathbf{b}, \mathbf{c})$. A closer look reveals that this is not the only way we can gradually construct a contradiction-free model in this case: we could have performed the assignments in a different order[5]. Of course, the only thing which would have been different would be the induced probabilities. That is, if one does the assignment in the following

---

[4]$\mathbb{P}(\mathbf{y})$ is induced by the assignment of $\mathbb{P}(\mathbf{x}, \mathbf{y})$.

[5]Yet, this is not always the case: suppose there are four RVs in the domain, namely, $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{d}$ and we would like to assign $\mathbb{P}(\mathbf{a}, \mathbf{b}), \mathbb{P}(\mathbf{b}, \mathbf{c})$, and $\mathbb{P}(\mathbf{c}, \mathbf{d})$. Performing the assignments in the order $(1) - \mathbb{P}(\mathbf{a}, \mathbf{b}), (2) - \mathbb{P}(\mathbf{b}, \mathbf{c}), (3) - \mathbb{P}(\mathbf{c}, \mathbf{d})$ would not introduce any inconsistencies, in contrast to using the order $(1) - \mathbb{P}(\mathbf{a}, \mathbf{b}), (2) - \mathbb{P}(\mathbf{c}, \mathbf{d}), (3) - \mathbb{P}(\mathbf{b}, \mathbf{c})$.

order: $(1) - \mathbb{P}(\mathbf{b}, \mathbf{d}), (2) - \mathbb{P}(\mathbf{a}, \mathbf{b}, \mathbf{c}), (3) - \mathbb{P}(\mathbf{b}, \mathbf{c}, \mathbf{e})$ then the first assignment of $\mathbb{P}(\mathbf{b}, \mathbf{d})$ will induce $\mathbb{P}(\mathbf{b})$ for the second assignment of $\mathbb{P}(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \mathbb{P}(\mathbf{b})\mathbb{P}(\mathbf{a}, \mathbf{c}|\mathbf{b})$ and the second assignment will induce $\mathbb{P}(\mathbf{b}, \mathbf{c})$ for the third assignment $\mathbb{P}(\mathbf{b}, \mathbf{c}, \mathbf{e}) = \mathbb{P}(\mathbf{b}, \mathbf{c})\mathbb{P}(\mathbf{e}|\mathbf{b}, \mathbf{c})$.

4 INFERENCE IN MCMS

In this section we consider *evidential* inference problems in multi-context settings. The objective is to evaluate (to the extent possible) a probability of the form $\mathbb{P}(\mathbf{O} = O|\mathbf{E} = E)$, called a *query*, where $\mathbf{O}$ and $\mathbf{E}$ are two mutually exclusive sets of RVs. The set $\mathbf{E}$ is the set of evidence variables and $\mathbf{O}$ is the set of RVs for which we are interested in knowing with what probability they take on the value O, upon the observation of $\mathbf{E} = E$. In multi-context settings, inference problems can be categorized into two broad classes:

- Intra-Contextual Inference Problems: For which the sets $\mathbf{E}$ and $\mathbf{O}$ both belong to the same context.

- Inter-Contextual Inference Problems: For which the sets $\mathbf{E}$ and $\mathbf{O}$ do not belong to a single context and, therefore, more than one context is involved in the inference problem.

In what follows, we will elaborate on these two cases.

4.1 INTRA-CONTEXTUAL INFERENCE PROBLEM

One advantage of MCMs is that, once an inference problem is found to be an intra-contextual inference problem, one can take advantage of the rich independence structure potentially governing the context to accomplish the task of inference in a computationally efficient way. For instance, if the probabilistic knowledge of a context is presented in a form of a BN, then one can benefit from a variety of exact or approximate methods already developed for BNs. For a comprehensive study of such methods the reader is referred to (Koller and Friedman 2009). Hence, it is of great interest to have contexts whose probabilistic knowledge can be represented in some form of a PGM with sufficiently rich independence structure for which inference problems can be solved in a computationally efficient way. For example, if the probabilistic knowledge of a context is to be modeled according to some BN, we would like that BN to be as sparsely connected as possible and enjoy low tree-width to ensure computational efficiency for the task of inference (Chandrasekaran, Srebro, and Harsha 2008).

4.2 INTER-CONTEXTUAL INFERENCE PROBLEM: INFERENCE GRAMMAR

In this section, we turn our attention to the task of inter-contextual inference. The RVs involved in the query for

the inter-contextual inference problem do not belong to a single context. For this reason, the answer to the query is inevitably in the form of an interval indicating a lower and upper bound for the query. Since $\mathbb{P}(E|O) + \mathbb{P}(\bar{E}|O) = 1$ we have $\mathbb{P}(E|O)_\uparrow = 1 - \mathbb{P}(\bar{E}|O)_\downarrow$. Therefore, we can focus our attention on the minimization problem (i.e., identifying a lower bound to the probability of interest) realizing that any maximization problem (i.e., identifying an upper bound to the probability of nterest) could be cast as a minimization problem and vice versa.

First, we are going to consider some simple queries which are posed to some example MCMs. These MCMs are depicted in Fig. 5(a-c). The goal here is to develop some insight as to which variables are indeed relevant and which are deemed irrelevant for a given query and the corresponding MCM.
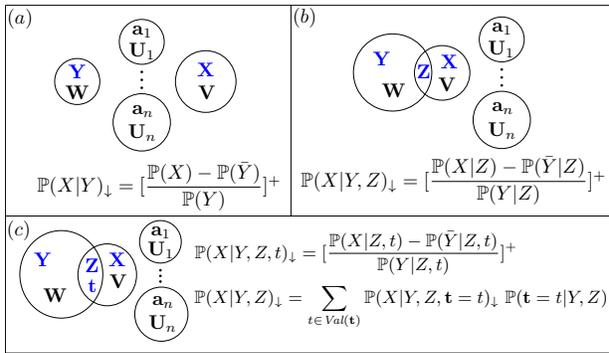

Figure 5: Sample inference rules given for some inter-contextual inference problems. The RVs involved in the query are shown in blue.

We begin by considering a simple case: the disjoint MCM shown in Fig. 5(a). The rule to evaluate $\mathbb{P}(X|Y)_\downarrow$ is also given in Fig. 5(a). Interestingly enough, the expression only requires the intra-contextual quantities $\mathbb{P}(X)$ and $\mathbb{P}(Y)$ and it does not depend on any other RV present in the domain. In other words, as far as $\mathbb{P}(X|Y)_\downarrow$ is concerned, the MCM shown in Fig. 5(a) is equivalent to a much simpler MCM: the one corresponding to having only two disjoint contexts described by $\mathbb{P}(\mathbf{X})$ and $\mathbb{P}(\mathbf{Y})$. Next, we take the MCM given in Fig. 5(b) where there is an overlap between the context containing $\mathbf{X}$ and the one containing $\mathbf{Y}$. The overlapping part consists of the random vector $\mathbf{Z}$. The rule to evaluate $\mathbb{P}(X|Y, Z)_\downarrow$ is given in Fig. 5(b). Now, consider the MCM shown in Fig. 5(c) where we have the same setting we had in previous case but a new random variable $\mathbf{t}$ is added in the overlapping region. Notice that the expression for $\mathbb{P}(X|Y, Z, t)_\downarrow$ given in Fig. 5(c) is the same expression given for $\mathbb{P}(X|Y, Z)_\downarrow$ in Fig. 5(b) with the substitution of Z, t instead of Z. That is, Z in Fig. 5(b) and Z, t in Fig. 5(c) are representing the same thing, namely, "all the variables in the overlapping region", and in that respect, they are ultimately the same. The rules are

very much like sentences in predicate logic for which variables merely serve as place-holders.

The derivation of the rules given in Fig. 5(a-c) is not presented here. However, using the proof presented in Sec. A-II of Appendix (to identify the relevant variables) and subsequently following the methodology outlined in Sec. A-III of Appendix (to visualize the partitions and reason out the extent they overlap) it should be straightforward to derive the presented rules.

The sample set of rules presented is by no means exhaustive, nonetheless, due to the idea of context transformation that will be discussed in Sec. 4.3, they can be applied to a wide range of interesting inter-contextual inference problems. We would like to clarify that our ultimate objective is *not* to compute and provide the complete set of rules that can answer all possible queries and for all possible MCMs, since simply, the set is infinite in size. What we need, therefore, is an algorithm, let us call it $\mathcal{I}^*$, that can provide the answer to the posed query being given an MCM as an input. The presented rules provide insights and hints to the nature of $\mathcal{I}^*$ which needs to be devised to ideally handle *any* arbitrary query posed to *any*[6] MCM. In a sense, we can get a glimpse of the nature of $\mathcal{I}^*$ through analyzing the presented rules. In other words, the derived rules serve as a lens through which one can study $\mathcal{I}^*$. In Sec. A-I of Appendix a simple version of $\mathcal{I}^*$ that can handle arbitrary MCMs is outlined.

The motivation behind giving this sample set of rules can now be summarized in the following.

1. To shed light on the general nature of a rule (which reflects on the nature of $\mathcal{I}^*$). More specifically, to illustrate that a rule enjoys two key properties, namely: (i) scale-invariance, (ii) resemblance to sentences in predicate logic, in that in both cases, variables are mere place-holders. For this resemblance we refer to $\mathcal{I}^*$ as *inference grammar*.

2. To demonstrate that a rule is telling us which intra-contextual quantities are essential and which are irrelevant for a particular inter-contextual query.

3. To emphasize the key property that a rule derived under a specific MCM remains valid for and can be applied to infinitely many other MCMs all of which are linked through the notions of nestedness and transformation; hence generalization is achieved.

4. To lay down the foundation of *transformation* and *nestedness* which both play crucial roles in understanding the underlying machinery behind $\mathcal{I}^*$.

---

[6]Although we believe that the MCMs generated through the generative process outlined in Sec. 3.2 are more cognitively plausible, nonetheless, from a pure mathematical point of view, it would be of interest to find an algorithm which could handle *any* MCM.

Next, we discuss another key property of the inference rules, namely, that of *scale-invariance*. Consider once again the case in Fig. 2. Now let us derive $\mathbb{P}(x_i|y)_\downarrow$, and $\mathbb{P}(X|y)_\downarrow$ where $\mathbf{X} \triangleq \mathbf{x}_{1:n}$. Using the rule given in Fig. 5(a), one arrives at the following results: $\mathbb{P}(x_i|y)_\downarrow = [\frac{\mathbb{P}(x_i)-\mathbb{P}(\bar{y})}{\mathbb{P}(y)}]^+$, and $\mathbb{P}(X|y)_\downarrow = [\frac{\mathbb{P}(X)-\mathbb{P}(\bar{y})}{\mathbb{P}(y)}]^+$. In other words, the expressions remain the same, regardless of the dimension of the quantity of interest, i.e., be it a single RV or be it a random vector comprised of many RVs. In this respect, once again, the inference rules resemble expressions in predicate logic. The intuition on the scale invariance is provided in Sec. A-III of Appendix.

It is worth noting that $\mathcal{I}^*$ formulates the inter-contextual inference problem as a Linear Programming (LP) optimization (cf. Sec. A-I of Appendix). The key issues to consider are: (i) what RVs have to be included in the LP, and (ii) the abstraction level $\mathcal{I}^*$ should choose to encode the RVs identified in step (i) for the LP, i.e., the parametrization of RVs identified in step (i) for the LP. In what follows, the concepts of nestedness and transformation are put forth. Once the two are introduced, one could apply a single rule (e.g., one in Fig. 5(a)) to a much larger number of MCMs; in fact to infinitely many MCMs.

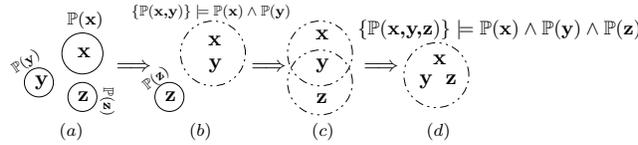### 4.3 INTER-CONTEXTUAL INFERENCE PROBLEM: NESTEDNESS AND TRANSFORMATION

Figure 6: Inter-Contextual Inference Problem: Transformation and hierarchical construct. As one proceeds from the left to the right, a more comprehensive knowledge of domain is assumed to be available, of course hypothetically.

The nested property, or *nestedness*, refers to the fact that every MCM can be considered as an element of a family of MCMs. That family contains all MCMs which through marginalization can produce the original MCM. In such a case we simply say that the nested property holds between the original MCM and the family. The process of going from the original MCM to one of the members of the family is referred to as *transformation*. For example, the MCM containing three contexts $\{\mathbf{x}\}$, $\{\mathbf{y}\}$, and $\{\mathbf{z}\}$ shown in Fig. 6(a) is a member of a family of MCMs containing two contexts $\{\mathbf{x}, \mathbf{y}\}$ and $\{\mathbf{z}\}$, shown in Fig. 6(b), one of which is associated to a *family* of JPDs over $\mathbf{x}$ and $\mathbf{y}$ (the dash-dotted circle in Fig. 6(b)) which, if marginalized, produces the same $\mathbb{P}(\mathbf{x})$ and $\mathbb{P}(\mathbf{y})$ in the original MCM (left-most MCM). Mathematically, the set of all

JPDs over RVs $\mathbf{x}$ and $\mathbf{y}$ which, if marginalized, produce specific marginal probability distributions $\mathbb{P}(\mathbf{x})$ and $\mathbb{P}(\mathbf{y})$ is denoted by $\{\mathbb{P}(\mathbf{x}, \mathbf{y})\} \models \mathbb{P}(\mathbf{x}) \wedge \mathbb{P}(\mathbf{y})$. The notion of the nested property enables us to look at one MCM as a subset of another larger MCM. The nested property, furthermore, enables one to sort MCMs in a hierarchical construct as illustrated in Fig. 6 where moving from the left to the right corresponds to moving from lower levels of hierarchy to higher levels.
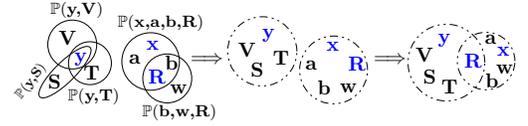

Figure 7: Transformation: Sample case.

To convey the idea, consider the case illustrated in Fig. 7. Suppose the query of interest is $\mathbb{P}(x|y, R)_\downarrow$. Then, one can first transform the original (left-most) MCM into the MCM shown in the middle, and subsequently into the right-most MCM. Hence, using the right-most MCM and the rule given in Fig. 5(b), one can write $\mathbb{P}(x|y, R)_\downarrow = [\frac{\mathbb{P}(x|R)-\mathbb{P}(\bar{y}|R)}{\mathbb{P}(y|R)}]^+ = [\frac{\mathbb{P}(x|R)-1+\mathbb{P}(y|R)}{\mathbb{P}(y|R)}]^+$. If we had the knowledge of $\mathbb{P}(y|R)$ then the expression given above would have been sufficient to derive $\mathbb{P}(x|y, R)_\downarrow$. However, since $\mathbb{P}(y|R)$ is *not* known, we need to go through one more step. This is precisely due to, and emphasizes, the fact that by working on the right-most MCM we implicitly presumed that we were equipped with more knowledge than we really had. Using the middle MCM and the rule given in Fig. 5(a), one can conclude $\mathbb{P}(y|R)_\downarrow = [\frac{\mathbb{P}(y)-\mathbb{P}(\bar{R})}{\mathbb{P}(R)}]^+$. Altogether[7], $\mathbb{P}(x|y, R)_\downarrow = ([\frac{\mathbb{P}(x|R)-1+\mathbb{P}(y|R)}{\mathbb{P}(y|R)}]^+)_\downarrow = [\frac{\mathbb{P}(x|R)-1+\mathbb{P}(y|R)_\downarrow}{\mathbb{P}(y|R)_\downarrow}]^+$. It is worth noting that the same rule would apply if instead of the random vector $\mathbf{R}$ we were dealing with the random variable $\mathbf{a}$, i.e., to find $\mathbb{P}(x|y, a)_\downarrow$ one could use the same expression given for $\mathbb{P}(x|y, R)_\downarrow$ by substituting a in place of R in all the expressions. Arguments of this kind are made possible due to the idea of transformation which enables us to analyze the transformed MCM (e.g., the middle one in Fig. 7) rather than the original MCM (the left-most one in Fig. 7). Furthermore, the concept of transformation highlights a key idea: if a piece of information (i.e., an intra-contextual quantity) is irrelevant in the transformed MCM for the posed query, it must have been irrelevant in the original MCM in the first place. This statement, once again, sheds light on what intra-contextual quantities are relevant or irrelevant to derive a posed inter-contextual query on a given MCM.

---

[7]This is due to the observation that for function $f(y) = (\frac{k+y}{y})$ when $k < 0$, $\min_{1 \geq y \geq t > 0} f(y) = (\frac{k+t}{t})$.

5   DISCUSSION

We will now discuss related work so as to build a connection between ours and previous attempts to incorporate partial probabilistic knowledge of a domain in the task of inference.

Attempting to combine Probabilistic Logic and BNs, the authors in (Andersen and Hooker 1990; 1994) formulate the inference problem as an optimization problem subject to non-linear constraints so as to incorporate the conditional independence relations embedded in the BN. However, in our proposed framework, the issue of dealing with conditional independence relations does not arise at all, because these relations are dealt with during the derivation process of intra-contextual probabilities.

The authors of (Hansen et al. 1995) point out that one could avoid non-linear optimization when the value for a conditional probability is at least imprecisely known. For example, the constraint $\mathbb{P}(a|b) = \mathbb{P}(a)$, if the value for $\mathbb{P}(a)$ is known either precisely or imprecisely within some interval $[\alpha, \beta]$, can be written as

$$\frac{\mathbb{P}(a,b)}{\mathbb{P}(b)} = \mathbb{P}(a) \in [\alpha, \beta] \Leftrightarrow \left\{ \begin{array}{l} \mathbb{P}(a,b) - \alpha\mathbb{P}(b) > 0, \\ \mathbb{P}(a,b) - \beta\mathbb{P}(b) < 0. \end{array} \right.$$

Hence, the independence $\mathbb{P}(a|b) = \mathbb{P}(a)$ can be formulated as a number of linear constraints. However, the main drawback of this approach is that encoding a conditional independence relation such as $\mathbb{P}(\mathbf{x}|\mathbf{y}, \mathbf{a}_1, \cdots, \mathbf{a}_n) = \mathbb{P}(\mathbf{x}|\mathbf{y})$ requires a number of linear equations that is exponential in n to be introduced into the optimization problem (Andersen and Hooker 1994).

Drawing on the idea of Context-Specific Independence (CSI) (Boutilier et al. 1996), the authors of (Geiger and Heckerman 1991) propose the Bayesian Multinet model which aims at taking advantage of the existing CSIs to perform inference, by modeling a single BN as multiple context-specific BNs. Translated into our multi-context setting, the Bayesian Multinet model corresponds to the case where the whole domain is modeled as a single BN, i.e., a single-context MCM, that can be decomposed into multiple BNs each being valid for a specific instantiation of some RVs in the domain.

The authors of (Thone, Guntzer, and Kiebling 1992) point out the same concerns which led us to propose MCM, namely: (i) If unverified (in)dependencies are imposed between the variables in the domain then implausible results may arise; (ii) PGMs require one to have complete probabilistic knowledge of a domain which may not be available. Motivated by these, (Thone, Guntzer, and Kiebling 1992) gives a collection of rules to carry out inference in a domain. Broadly speaking, this work is similar to ours in spirit with the main distinction being the level of abstraction chosen to perform inference. In (Thone, Guntzer, and

Kiebling 1992) inference is performed in a very local and rule-based fashion and conditional independence relations are dealt with directly which complicates the task at hand; a task which is futile when it comes to dealing with domains of many variables. In our case, by introducing the notion of context and encoding conditional independence relations within contexts we avoid having to contemplate the intra-contextual inference problem and leave this task for the corresponding context. This way, we can take advantage of the possibly rich independence structure governing the context and carry out the intra-contextual inference problem in a computationally efficient manner.

Finally, let us discuss some interesting aspects of the proposed model.

The degree of belief is encoded mathematically in the form of a probability distribution over the variables contained within the context. Furthermore, in the process of partial belief formation (which leads to the formation of contexts) the reasoner is ignorant as to how various contexts probabilistically interact (are related), except that, some contexts may in fact share a number of variables in between and hence overlap. Later on, in the process of the derivation of the query posed to the reasoner, this ignorance manifests in the uncertainty region represented by the min/max values for the inter-contextual query of interest. In other words, if the reasoner incurs ignorance as to the (in)dependency structure governing the variables present in the domain, then later on, in the process of derivation of the posed query, the reasoner has to pay the price by merely arriving at a *probability interval* rather than a point probability as an answer to the query of interest. Yet, the knowledge of the underlying dependency structure is a fundamental knowledge whose availability to the reasoner should *not* be postulated as an inevitability but as an advantaged position.

The evolutionary process of MCM does not enforce a specific gradual expansion path, for the claim of MCM is merely that *any partial belief formation as to the domain can be modeled in the framework depicted by MCM*. That is, the reasoner may arrive at different MCMs, depending on the order in which the reasoner encounters different concepts and also depending on her background knowledge as to the nature of the potential connections between a collection of variables. Simply put, the order according to which the reasoner comes about knowing the concepts or propositions of the domain does matter (cf. the discussion on the order of belief formation in Sec. 3.2).

MCM enables one to carry out inference without having to commit to any unjustified or uncertain independence assumptions. In light of this, contexts symbolize the regions of the domain over which an (in)dependence structure is presumed and hence, the growth and merging of contexts indicates the formation of new (in)dependence structures over some parts of the domain which previously were un-

structured. In short, MCM is meant to be invoked in circumstances where the observations and the a priori knowledge combined are not sufficient for the reasoner to form the full JPD over all of the domain variables and yet, quite crucially, the reasoner is reluctant to submit to any unjustified assumptions to compensate for such inadequacy of knowledge.

6 CONCLUSION

In an attempt to establish a middle ground between Bayesian Logic and Probabilistic Logic (Andersen and Hooker 1990; 1994), on one side, and PGMs[8] on the other, we proposed the Multi-Context Model to represent the state of partial knowledge regarding a domain. The generative process for the gradual construction of contradiction-free MCMs was discussed. The task of Inference for MCM was studied and, along the path, the notions of inference grammar, nestedness, and transformation were introduced. A short version of $\mathcal{I}^*$ without the scale-invariance property was provided in Appendix. It is worth noting that scale-invariance property can be achieved with a minor change to the last step of the proposed algorithm.

APPENDIX

A-I $\mathcal{I}^*_{non-scale}$: A short version of $\mathcal{I}^*$ without scale-invariance property

$\mathcal{I}^*$ aims at minimally parameterizing the information contained in an MCM so that the posed inter-contextual query can be stated as an LP with the fewest number of parameters. As pointed out earlier in Sec. 4.2, $\mathcal{I}^*$ has to decide on the following: (i) what RVs have to be included in the LP, and (ii) the abstraction level required to minimally encode the information on the RVs identified in step (i) for the LP, in our case, the parametrization of the identified RVs.

In what follows, a simple algorithm, $\mathcal{I}^*_{non-scale}$, is sketched which only performs (i) and ignores (ii). In other words, $\mathcal{I}^*_{non-scale}$ identifies the relevant RVs needed to derive the *exact* lower/upper bound for the inter-contextual query, however, it does not aim at minimally encoding them into the LP[9]. $\mathcal{I}^*_{non-scale}$ consists of three steps:

(1) Identify all the RVs involved in the posed query (e.g., in $P(X|Y, z)$ these are the random vector X, random vector Y and RV z).

(2a) If any two of the already identified RVs belong to two

---

[8] For instance, Bayesian Networks (Pearl 1986), Markov Networks (Koller and Friedman 2009), and Chain Graphs (Buntine 1995).

[9] To read more on this, the reader is referred to the discussion on scale-invariance property in Sec. 4.2 and Sec. A-III of Appendix.

overlapping contexts, identify all the *overlapping* RVs between these two contexts (e.g., in Fig. 5(b) and for the query $P(X|Y)$ for which step (1) would identify X and Y, random vector Z in the overlapping region must be identified as well).

(2b) If any two of the already identified RVs belong to two contexts connected through a chain of overlapping contexts: identify all the RVs contained in all the *overlapping* regions of the chain of contexts.

(3) Parameterize only the identified RVs in steps (1), (2a), and (2b) (remove all the other RVs from the MCM—there is no need to encode the information on any other RVs not identified in steps (1), (2a), and (2b)).

∎

It should be noted that whether the posed query involves minimization or maximization does not affect which RVs need to be identified by $\mathcal{I}^*_{non-scale}$. Finally, It is worth noting that with a minor modification to step (3) of $\mathcal{I}^*_{non-scale}$, the scale-invariance property could be achieved. The modification has to do with the question of how to *minimally* encode the information on each RV identified in steps (1), (2a), and (2b) of $\mathcal{I}^*_{non-scale}$.

To demonstrate the operation of $\mathcal{I}^*_{non-scale}$ on a more complicated MCM that involves loops, consider the following example sketched in Fig. 8(a). The query of interest is $\mathbb{P}(X|Y)_\downarrow$.
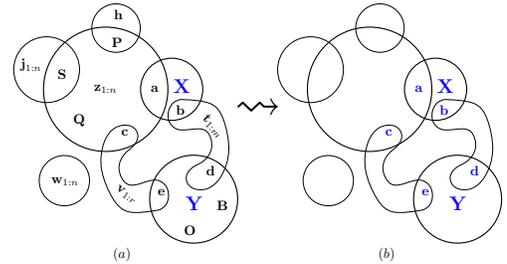

Figure 8: (a) Sample MCM. The RVs involved in the posed query are depicted in blue. (b) In Step (1) **X** and **Y** are identified; in step (2b) the RVs **b**, **d** as well as **a**, **c**, and **e** are identified. According to step (3) of $\mathcal{I}^*_{non-scale}$ all of the information as to the RVs **X**, **Y**, **b**, **d**, **a**, **c**, and **e** has to be stated as an LP to derive the query.

Next, we are going to sketch the proof for $\mathcal{I}^*_{non-scale}$. Let us first state the claim formally and then provide the proof.

A-II Proof for $\mathcal{I}^*_{non-scale}$:

Lemma: *Given a posed query and an MCM, if all the information on the RVs identified in steps (1) to (2b) of $\mathcal{I}^*_{non-scale}$ is stated and then solved as an LP, the exact solution (i.e., a min or max) can be derived for the posed*

query; all the remaining information available in the MCM is deemed irrelevant to the derivation of the query, hence the sufficiency.

Proof: Our proof is constructive. In the proof we entertain two ideas, namely (i) the idea of generative process and, particularly, that of *conditioning* also used in Sec. 3.2, and (ii) the notion we refer to as the *locality of information*. Suppose that all the RVs discussed in steps (1) to (2b) of $\mathcal{I}^*_{non-scale}$ are identified. The key insight is that the information on how the remaining RVs probabilistically interact with each other is completely local in nature and, therefore, irrelevant to the derivation of the posed query. To see this, one can start off with the identified RVs and then in a gradual fashion add on[10] the rest of the RVs (through the idea of conditioning discussed in Sec. 3.2). Quite crucially, this very process of adding the non-identified RVs to the model can be done completely in a local fashion, i.e., without imposing any constraints on how the identified RVs probabilistically interact. The mere fact that those RVs can be added into the model: (i) subsequent to the identified ones, and (ii) without inducing any sort of constraints on the identified ones, deems them irrelevant to the derivation of the query. ∎

A-III Scale-Invariance Property: Intuition

Here, we will provide a proof for the example on scale-invariance property given in Sec. 4.2. Although the proof is provided for a special query, the methodology used in the proof provides an insightful way of *visualizing* an inference problem. The idea behind the proof is very simple and related to visualizing the connection of a RV to the underlying sample space using Venn diagrams. Without loss of generality, we assume that all the RVs present in the domain are binary[11]. Random vector $\mathbf{X} = \mathbf{x}_{1:n}$ partitions the sample space Ω into 2^n disjoint regions each of which corresponds to a realization of $\mathbf{X}$. If each realization of the random vector $\mathbf{x}_{1:n}$ corresponds to a binary number (i.e., binary-coding the realizations), then one can conclude $Val(\mathbf{X}) = \{0, 1, \cdots, 2^n - 1\}$. Let us index the partitions by their corresponding realization of $\mathbf{X}$. An illustrative example of an induced partitioning of the sample space Ω due to random vector $\mathbf{X} = \mathbf{x}_{1:n}$ is depicted in Fig. 9(a), and a partitioning induced by RVs $\mathbf{y}$ and $\mathbf{z}$ is sketched in Fig. 9(b). We note that the mere knowledge of the distribution function of a random quantity does not provide one with the knowledge of the underlying partitions. For this particular example, since the JPD over $\mathbf{X}, \mathbf{y}, \mathbf{z}$ is not available, the knowledge of how the partitions induced by $\mathbf{y}, \mathbf{z}$ (Fig. 9(b)) and the ones induced by $\mathbf{X}$ (Fig. 9(a)) interact, i.e., to what extent they overlap,

remains unspecified. Therefore, since $\mathbb{P}(X|y) = \frac{\mathbb{P}(X,y)}{\mathbb{P}(y)}$, to minimize (maximize) $\mathbb{P}(X|y)$, the quantity $\mathbb{P}(X, y)$ has to be minimized (maximized). Pictorially, the minimization (maximization) of $\mathbb{P}(X, y)$ corresponds to the minimization (maximization) of the overlap between the partitions corresponding to the events $\{\mathbf{X} = X\}$ and $\{\mathbf{y} = y\}$; hence, very simply, $\mathbb{P}(X,y)_\downarrow = [\mathbb{P}(X) + \mathbb{P}(y) - 1]^+$ and $\mathbb{P}(X,y)_\uparrow = \min\{\mathbb{P}(X), \mathbb{P}(y)\}$. The key point, which yields the scale-invariance property, is that to derive the minimum (maximum) overlap between the partitions corresponding to the events $\{\mathbf{X} = X\}$ and $\{\mathbf{y} = y\}$ *the information as to how the other partitions—corresponding to the other realizations of the present RVs in the model—interact with one another neither needs to be known nor to be encoded into the LP*; a fact which results in not requiring to encode the information as to the other realizations. Hence the only pieces of information that are required to be encoded and then solved as an LP are $\mathbb{P}(X)$ and $\mathbb{P}(y)$. The same line of reasoning could be adopted for $\mathbb{P}(x_i|y)$. The idea of scale-invariance, therefore, aims to avoid the encoding of the information as to the partitions induced on Ω which are yet deemed to be irrelevant to the derivation of the posed query; hence one needs to encode solely the relevant ones into the LP.
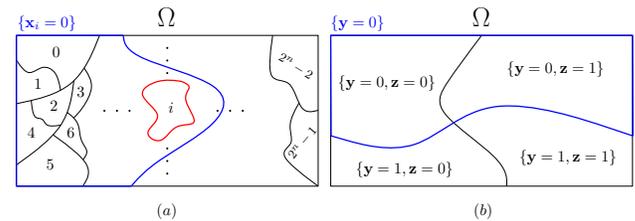

Figure 9: Sample Space: (a) Partitioning induced on Ω due to $\mathbf{X} = \mathbf{x}_{1:n}$. The blue region corresponds to the partition associated to the event $\{\mathbf{x}_i = 0\}$ and the red one to that of $\{\mathbf{X} = i\}$ where $i \in Val(\mathbf{X})$. (b) Partitioning induced on Ω due to RVs $\mathbf{y}$ and $\mathbf{z}$. The blue region corresponds to the partition associated to the event $\{\mathbf{y} = 0\}$.

---

[10]This is based on the fundamental property that a JPD can be expanded using the chain rule of probability in an arbitrary order.

[11]The generalization of the argument to non-binary RVs is straightforward.

References

Andersen, K. A., and Hooker, J. N. 1990. Probabilistic logic for belief nets. In *International Congress of Cybernetics and Systems*. New York City.

Andersen, K. A., and Hooker, J. N. 1994. Bayesian logic. *Decision Support Systems* 11(2):191–210.

Boutilier, C.; Friedman, N.; Goldszmidt, M.; and Koller, D. 1996. Context-specific independence in bayesian networks. 115–123. Morgan Kaufmann Publishers Inc.

Buntine, W. L. 1995. Chain graphs for learning. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, 46–54. Morgan Kaufmann Publishers Inc.

Chalmers, A. F. 1976. *What is this thing called science?* Hackett Publishing.

Chandrasekaran, V.; Srebro, N.; and Harsha, P. 2008. Complexity of inference in graphical models. *Proceedings of the 24th Conference in Uncertainty in Articial Intelligence* 70–78.

Geiger, D., and Heckerman, D. 1991. Advances in probabilistic reasoning. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, 118–126. Morgan Kaufmann Publishers Inc.

Hansen, P.; Jaumard, B.; Nguetse, G. D.; and Aragao, M. P. D. 1995. *Models and algorithms for probabilistic and Bayesian logic*. Citeseer.

Koller, D., and Friedman, N. 2009. *Probabilistic graphical models: principles and techniques*. MIT press.

Pearl, J. 1985. Bayesian networks: a model of self-activated memory for evidential reasoning. *in Proceedings of the Cognitive Science Society* 329–334.

Pearl, J. 1986. Fusion, propagation, and structuring in belief networks. *Artificial intelligence* 29(3):241–288.

Pearl, J. 1990. Reasoning with belief functions: An analysis of compatibility. *International Journal of Approximate Reasoning* 4(5):363–389.

Thone, H.; Guntzer, U.; and Kiebling, W. 1992. Towards precision of probabilistic bounds propagation. In *Uncertainty in Artificial Intelligence*, 315–322.