Statistical Relational Artificial Intelligence with Relative Frequencies: A Contribution to Modelling and Transfer Learning across Domain Sizes

Felix Weitkämper
Institut für Informatik, LMU München, Munich, Germany
felix.weitkaemper@lmu.de

Abstract

Dependencies on the relative frequency of a state in the domain are common when modelling probabilistic dependencies on relational data. For instance, the likelihood of a school closure during an epidemic might depend on the proportion of infected pupils exceeding a threshold, say 1%. Continuous dependencies are also common: for instance; the likelihood of any one mosquito bite transmitting an illness depends on the proportion of carrier mosquitoes. Current approaches usually only consider probabilities over possible worlds rather than over domain elements themselves. We introduce two formalisms that explicitly incorporate relative frequencies into statistical relational artificial intelligence. The first formalism, Lifted Bayesian Networks for Conditional Probability Logic, expresses discrete dependencies on probabilistic data. The second formalism, Functional Lifted Bayesian Networks, expresses continuous dependencies. Incorporating relative frequencies is not only beneficial to modelling; it also provides a more rigorous approach to learning problems where training and test or application domains have different sizes. To this end, we provide a representation of the asymptotic probability distributions induced by the two formalisms on domains of increasing sizes. Since that representation has well-understood scaling behaviour across domain sizes, it can be used to estimate parameters for a large domain consistently from randomly sampled subpopulations.

1 Introduction

Consider the different flavour of the following two statements: “ 1% of the population are suffering from the disease”, which is a statement about the relative frequency of an illness in the population; and “Considering his symptoms, the likelihood that this patient is suffering from the disease is 20%”, which is a statement about the degree of confirmation of the assertion that a particular patient has this illness, given the available evidence.

This distinction has first been methodically investigated by Carnap (1950), who distinguished two concepts of probability, the “degree of confirmation”, which he calls “probability$_1$”, and the “relative frequency (in the long run)”, which he calls “probability$_2$”. Carnap goes on to formalise probability$_1$ using a probability measure defined over so-called *state descriptions*, which we can identify as *possible worlds* in more modern terminology. Probability$_2$ on the other hand is interpreted by the uniform measure on a given domain set itself.

Forty years later, in his seminal paper on the analysis of first-order logics of probability (1990), Halpern divided approaches to formalising probability in a relational setting along very similar lines. Halpern refers to logics encoding relative frequencies as Type I logics, while referring to logics encoding degrees of belief as Type II logics [1]. As a distinct category, Halpern also considers logics that combine both by expressing a degree of belief in a statement that mentions relative frequencies. He refers to those as Type III logics. Type III logics can express compound statements such as “With a likelihood of at least 10%, more than 60% of the population will have been ill by the end of the year.”

In our contribution, we outline the Halpern types corresponding to the prevalent formalisms of statistical relational artificial intelligence and investigate the benefits of a Type III semantics to this context. They include appropriateness to the intended meaning of queries and a better grasp of extrapolation behaviour, which enables transfer learning across domain sizes. We then discuss two concrete Type III formalisms, the Lifted Bayesian networks for Conditional probability logic (LBN-CPL) introduced by Koponen (2020) and Functional Lifted Bayesian Networks (FLBN), proposed here. We explore the expressiveness of those formalisms and give a formal account of their extrapolation behaviour. After briefly summarising the learning algorithms that are available or adaptable to LBN-CPL and FLBN, we give a formal account of their extrapolation behaviour with increasing domain size and explain how this addresses the transfer learning problem posed in the introduction.

The main original contributions of this article are:

1. We introduce and evaluate FLBN as a statistical relational Halpern Type III formalism for continuous

---

[1] Slightly unfortunately, this terminology is reverse to that of Carnap mentioned above

dependencies

2. We propose and evaluate the use of LBN-CPL as a statistical relational Halpern Type III formalism for continuous dependencies
3. We characterise the asymptotic behaviour of FLBN on domains of increasing sizes.
4. We outline an asymptotically consistent approach to transfer learning across domain sizes using the asymptotic representations.

1.1 Current formalisms in statistical relational AI

As Muggleton and Chen (2008) and Schulte (2012) have noted, the vast majority of statistical relational frameworks in use today are of Halpern Type II - they allocate a probability to each possible world.

We briefly outline the situation for Markov Logic Networks (MLN), Probabilistic Logic Programming under the distribution semantics (PLP) and approaches based on lifting Bayesian networks.

In their simplest form, MLNs (Richardson and Domingos 2006) are given by a set of first-order formulas φ_i in a signature σ annotated with real-valued weights w_i, as well as a domain D. Then a probability measure over the set of all possible worlds $\mathfrak{X}$ on D (i. e. σ-structures with domain D) is defined by setting

$$\mathcal{P}(X = \mathfrak{X}) := \frac{1}{Z} \exp\left(\sum_i w_i n_i(\mathfrak{X})\right)$$

where $n_i(\mathfrak{X})$ is the number of true groundings of φ_i in $\mathfrak{X}$ and Z is a normalisation constant ensuring that the probabilities of all possible worlds sum to 1.

The distribution semantics for PLP (Riguzzi and Swift 2018) is given by a stratified logic program over independently distributed probabilistic ground facts on a domain D. More precisely, let $\rho \subseteq \sigma$ be signatures and let $\alpha_i \in [0, 1]$ for atoms $R_i(x)$ from ρ. Let Π be a stratified Datalog program with extensional vocabulary ρ and intensional vocabulary $\sigma \setminus \rho$. Then every ρ-structure $\mathfrak{Y}$ on D induces a σ-structure $\mathfrak{X}$ on D obtained by evaluating Π with input $\mathfrak{Y}$. In this way, Π lifts the probability distribution on ρ-structures given by making independent choices $R_i(x)$ of atoms from ρ with probabilities α_i,

$$\mathcal{P}(X = \mathfrak{Y}) := \left(\prod_{a \in D, R_i \in \rho, \mathfrak{Y} \models R_i(a)} \alpha_i\right)\left(\prod_{a \in D, R_i \in \rho, \mathfrak{Y} \models \neg R_i(a)} (1 - \alpha_i)\right)$$

to a probability distribution on σ-structures with domain D.

Approaches based on lifting Bayesian networks, such as Relational Bayesian Networks (RBN) (Jaeger 2002), Bayesian Logic Programs (BLP) (Muggleton and Chen 2008) and Relational Logistic Regression (RLR) (Kazemi et al. 2014), provide a template for a Bayesian network on any given domain D, with a node for every possible ground atom $R(a)$, $a \in D$. The probability of every possible world is then defined in the manner usual for Bayesian networks. We briefly sketch the RLR formalism as an example: Here, the nodes of the DAG are given by atoms of σ, and every node $Q(\vec{x})$ is annotated with a list $(\varphi_i, w_i)_i$ of formulas whose relations are taken from the parents of $Q(\vec{a})$ and real-values weights w_i. The probability of $Q(\vec{a})$ given a grounding of its parents is as follows:

$$\mathcal{P}(Q(\vec{a})) := \operatorname{sigmoid}\left(\sum_i w_i n_i\right)$$

where n_i is the number of true groundings of the formula φ_i.

As a generalisation of stochastic grammars, Stochastic Logic Programs (SLP) (Muggleton and Chen 2008) are very different to the approaches above. Rather than providing a probability distribution over possible worlds, they define a distribution over derivations of a goal in a logic programming setting. Since deriving a goal equates to querying a Herbrand base, this can be seen as defining a distribution within that model. Therefore, SLPs can be classified as Type I.

More explicitly of Type I is the class-based semantics for parametrised Bayesian networks suggested by Schulte et al. (2014). Syntactically, they are similar to the template Bayesian networks mentioned above, but probabilities are defined without grounding to any specified domain. Instead, they are interpreted as arising from a random choice of substitutions that instantiate the template nodes.

To the best of our knowledge, no statistical relational approach has yet been identified as Type III of the classification. However, Jaeger (2002) has provided for a 'mean' aggregation function for RBN and Weitkämper (2021) has investigated scaled Domain-size Aware RLR (DA-RLR) in which parameters are scaled with domain size. Both formalisms induce a dependency on relative frequency (Type I probability) of domain atoms, and we will see in Subsection 3.2 below how those approaches are subsumed by our Type III framework.

1.2 Queries relating to degrees of belief vs relative frequency

Picking up the thread of disease modelling, we outline how probabilistic models and queries fit into the context of probability types.

Example 1. For a domain of people and a graph of connections between them, consider a model that represents how connected individuals infect each other. A Type II query would ask "what is the likelihood that a given individual is infected at time t (possibly given evidence)".

This is clearly an interesting problem on this domain. However, a main focus of epidemiologic modelling are adaptive interventions: A trigger is set (such as

"1% of the population are infected") and then some intervention is performed (such as "schools are closed") as soon as that trigger is reached.

Such trigger conditions that refer to relative frequencies are very common (see Bisset et al. (2014, Table II) for further examples). This naturally leads to Type III queries, in which the likelihood of a certain frequency event is addressed: "How likely is it that 1% of the population will be infected within the next four weeks?"

We introduce two different Type III formalisms, both of which incorporate Type I expressions into a Type II framework of lifted Bayesian networks. They are distinguished by the type of dependencies that they are designed to model. LBN-CPL is built around the Type I language 'Conditional Probability Logic', in which conditions such as "At least 5% of pupils in school s are diagnosed" can be expressed. Probabilities of relations or propositions at a child node would then depend on which condition is satisfied: Often, however, dependencies are not discrete - instead, they are of the form "the higher (or lower) the proportion of R, the more likely is Q". An example would be a mosquito-borne disease, in which the transmission risk of a single bite is directly proportional to the relative frequency of disease carriers in the mosquito population. Therefore we also introduce FLBN, where the probability of a child relation is a continuous function of the relative frequencies of parent nodes.

1.3 Transfer learning and Extrapolation

We will see that systematically using Type I probabilities within an outer framework of lifted Bayesian networks also addresses a pertinent problem in parameter learning for statistical relational representations: discrepancies in domain size between training and test or deployment sets. Such a discrepancy could occur in different settings. On one hand, it could be a deliberate choice since learning can be considerably more expensive than inference (which is known to be NP-hard in general (Dalvi and Suciu 2012)). Therefore, sampling small subsets of a complete dataset and then training the parameters on the sampled subsets could be much more efficient than learning the parameters on the entire set. This is recommended by the authors of the MLN system Tuffy (Doan et al. 2011), for instance. On the other hand, the size of the test set might be variable or unknown at training time.

It is well-known, however, that in general the parameters that are optimal on a randomly sampled subset are not optimal on the larger set itself.

Example 2. Consider the typical example of a relational logistic regression with two unary relations R and Q (cf. (Poole et al. 2014)), and an underlying DAG $R(x) \longrightarrow Q(y)$. For any $b \in D$, the probability of $Q(b)$ is given by $\text{sigmoid}(w * |a \in D | R(a)|)$, where w is a parameter to be learned from data. Now consider a training set of domain size 1000 in which 100 elements each satisfy R and Q. Now assume that we sample a subset of 100 elements in which 10 elements each satisfy R and Q. The optimal parameter on that subset would be a w for which $\text{sigmoid}(w * 10) = 10/100$, which turns out to be around -0.21. On the original set, however, the optimal parameter satisfies $\text{sigmoid}(w * 100) = 100/1000$, which is around -0.021. Indeed, if we would transfer the original parameter to the larger set, it would predict a probability for $Q(y)$ of less than 10^{-9}!

Jaeger and Schulte (2018) showed that for certain projective families of distributions such a sampling approach provides a statistically consistent estimate of the optimal parameters on the larger set. However, projectivity is a very limiting condition; in fact, the projective fragments of common statistical relational frameworks that are isolated by Jaeger and Schulte (2018) are essentially propositional and cannot model any interaction between an individual and the population at large. For example, to make the relational logistic regression above projective, the probability of $Q(a)$ must only depend on whether $R(a)$ is true, and not on any other elements of the domain. We will show that despite their larger expressivity, Type III frameworks can be meaningfully approximated by projective distributions on large domains, allowing us to leverage the statistical consistency results for projective families of distributions.

2 Introducing Type III formalisms

We propose two Type III formalisms for statistical relational AI. We establish the Type I language of conditional probability logic (CPL) and then extend it to a Type III framework using lifted Bayesian networks, following Koponen (2020). We then propose a second framework, able to express continuous rather than discrete dependencies on Type I probabilities.

As CPL is an extension of classical first-order logic, we will assume the syntax and semantics of first-order logic to be known (see e. g. Ebbingaus et al.(1994)).

2.1 Conditional Probability Logic

CPL formulas over σ are defined inductively as follows (Koponen 2020, Definition 3.1):

Definition 1. Let σ be a (possibly multi-sorted) relational signature (with equality). Then the set of *conditional probability formulas* over σ is defined as follows:

1. For every relation symbol R in σ of arity n and appropriate terms (i. e. variables or constants of the correct sorts) $t_1, \ldots, t_n$, $R(t_1, \ldots, t_n)$ is a conditional probability formula.
2. If ϕ, ψ are conditional probability formulas, then $\neg\varphi, \varphi \wedge \psi, \varphi \vee \psi$ and $\varphi \rightarrow \psi$ are also conditional probability formulas.

3. If φ is a conditional probability formula and x is a variable, then $\forall_x \varphi$ and $\exists_x \varphi$ are also conditional probability formulas.
4. For any $r \in \mathbb{R}_+$, conditional probability formulas φ, ψ, θ and τ, and a tuple of distinct variables $\vec{y}$, the following are conditional probability formulas:

$$r + \|\varphi|\psi\|_{\vec{y}} \geq \|\theta|\tau\|_{\vec{y}}$$

$$\|\varphi|\psi\|_{\vec{y}} \geq \|\theta|\tau\|_{\vec{y}} + r$$

In ordinary first-order logic, a variable is *bound* if it is in the range of an existential or universal quantifier. In conditional probability logic, a variable is also bound if it is in the range of a construction of the form $\|\varphi|\psi\|_{\vec{y}} \geq \|\theta|\tau\|_{\vec{y}} + r$ or $\|\varphi|\psi\|_{\vec{y}} + r \geq \|\theta|\tau\|_{\vec{y}}$.

The semantics for CPL are an extension of the ordinary semantics of first-order logic. Intuitively, the new constructor $\|\varphi|\psi\|_{\vec{y}}$ should be read "The conditional (Type-I-)probability that $\varphi(\vec{y})$ holds given that $\psi(\vec{y})$ is known to hold."

Koponen (2020) only defines the semantics on finite structures, where the Type I probability measure on the domain is given by the counting measure. This is also the setting most relevant to statistical relational learning and reasoning.

Definition 2. Let σ be a relational signature and let $\mathfrak{X}$ be a finite σ-structure (in the sense of ordinary first-order logic with equality) on domain D. We define what it means for a conditional probability formula φ to *hold in* $\mathfrak{X}$ with respect to any sort-respecting interpretation of variables ι, written as $\mathfrak{X} \models_\iota \varphi$. Note that Clauses 1 through 3 of Definition 1 are taken from ordinary first-order logic. Therefore, we can copy the recursive definition of atomic formulas and connectives directly from the corresponding clauses for first-order logic. So assume that $\models_\iota$ has been defined for φ, ψ, θ and τ. Let $\iota_{\vec{y}:\vec{b}}$ be the variable interpretation obtained from ι by mapping $\vec{y}$ to $\vec{b}$ and otherwise following ι. Furthermore, for any term t let D_t be the domain of the sort of t. We then define $|\varphi|_{\vec{y},\iota}$ to be the cardinality of $\left\{ \vec{b} \in \prod_{y \in \vec{y}} D_y | \mathfrak{X} \models_{\iota_{\vec{b}}} \varphi \right\}$.

We set $\mathfrak{X} \models_\iota r + \|\varphi|\psi\|_{\vec{y}} \geq \|\theta|\tau\|_{\vec{y}}$ if and only if $|\tau|_{\vec{y},\iota} > 0$, $|\psi|_{\vec{y},\iota} > 0$ and $r + \frac{|\varphi \wedge \psi|_{\vec{y},\iota}}{|\psi|_{\vec{y},\iota}} \geq \frac{|\theta \wedge \tau|_{\vec{y},\iota}}{|\tau|_{\vec{y},\iota}}$.

Analogously, $\mathfrak{X} \models_\iota \|\varphi|\psi\|_{\vec{y}} \geq \|\theta|\tau\|_{\vec{y}} + r$ if and only if $|\tau|_{\vec{y},\iota} > 0$, $|\psi|_{\vec{y},\iota} > 0$ and $\frac{|\varphi \wedge \psi|_{\vec{y},\iota}}{|\psi|_{\vec{y},\iota}} \geq \frac{|\theta \wedge \tau|_{\vec{y},\iota}}{|\tau|_{\vec{y},\iota}} + r$.

We introduce some intutitive shorthands:

Definition 3. We write $\|\varphi\|_{\vec{y}}$ for $\|\varphi|y_1 = y_1\|_{\vec{y}}$, expressing the unconditional probability of $\varphi(\vec{y})$, and we write $r \geq \|\theta|\tau\|_{\vec{y}}$ and $\|\varphi|\psi\|_{\vec{y}} \geq r$ for $r + \|\neg(y_1 = y_1)|y_1 = y_1\|_{\vec{y}} \geq \|\theta|\tau\|_{\vec{y}}$ and $\|\varphi|\psi\|_{\vec{y}} \geq \|\neg(y_1 = y_1)|y_1 = y_1\|_{\vec{y}} + r$ respectively. We will refer to the set of all conditional probability formulas over a relational signature σ as CPL(σ).

CPL can express the trigger functions in epidemic modelling that we had mentioned in the previous section: for instance, "at least 1% of domain individuals are infected" is simply $\|\text{Infected}(x)\|_x \geq 0.01$. Using conditional probabilities, more complex relationships can also be expressed: "at least 5% of pupils at school s are infected" can be expressed as $\|\text{Infected}(x)|\text{Pupil}(x,s)\|_x \geq 0.05$. As an example utilising the full syntax including nested probability quantifiers, we can even extend this to "School s is at least at the median among schools in area a by proportion of infected pupils":

$$\begin{aligned} 0.5 \geq \Big\| \|\text{Infected}(x)|\text{Pupil}(x,s)\|_x & \\ \geq \|\text{Infected}(x)|\text{Pupil}(x,y)\|_x & \Big| \text{Located}(y,a) \Big\|_y \end{aligned}$$

2.2 Lifted Bayesian Networks for Conditional Probability Logic

To extend this Type I logic to a Type III representation and to integrate it with the independence assumptions from Bayesian Networks, we follow Koponen (2020) in introducing *Lifted Bayesian Networks.*

Definition 4. A *lifted Bayesian network for conditional probability logic (LBN-CPL)* over a relational signature σ consists of the following:

1. An acyclic directed graph (DAG) G with node set σ.
2. For each $R \in \sigma$ a set of formulas $\chi_{R,i}(\vec{x}) \in$ CPL(par(R)), where (par(R) is the signature of the G-parents of R, $|\vec{x}|$ is a sort-appropriate tuple of the correct length for R, such that $\forall \vec{x} \big(\bigvee_{i=1}^{\nu_R} \chi_{R,i}(\vec{x}) \big)$ is valid (i.e. true in all par(R)-structures) and if $i \neq j$ then $\exists \vec{x} \big(\chi_{R,i}(\vec{x}) \wedge \chi_{R,j}(\vec{x}) \big)$ is unsatisfiable. Such a set of $\chi_{R,i}$ is called a *partition.*
3. For each $R \in \sigma$ and each associated formula, a number denoted $\mu(R \mid \chi_{R,i})$ (or $\mu(R(\vec{x}) \mid \chi_{R,i}(\vec{x}))$) in the interval $[0,1]$.

The semantics of lifted Bayesian networks are defined by grounding to a Bayesian network with respect to a given domain. So let D be a finite domain. Then we can view a σ-structure with domain D as a choice of truth value for each $R(\vec{a})$, where R is a relation symbol in σ and $\vec{a}$ is a tuple of elements of D of the right length and the right sorts for R. Therefore, defining a probability distribution over the space of possible σ-structures with domain D is equivalent to defining a joint probability distribution over the $R(\vec{a})$, viewed as binary random variables.

Definition 5. Consider an LBN-CPL $\mathfrak{G}$ and a finite domain D. Then the probability distribution induced by $\mathfrak{G}$ on the set of σ-structures with domain D is given by the following Bayesian network: The nodes are given by $R(\vec{a})$, where R is a relation symbol in σ and $\vec{a}$ is a tuple of elements of D of the right length and the right sorts for R. There is an edge between two nodes $R_1(\vec{a})$ and $R_2(\vec{b})$ if there is an edge between R_1 and R_2 in the

DAG G underlying $\mathfrak{G}$. It remains to define a probability table for every node $R(\vec{a})$: Given a choice of values for $P(\vec{b})$ for all $P \in \mathrm{par}(R)$ and appropriate tuples $\vec{b}$ from D, the probability of $R(\vec{a})$ is set as $\mu(R \mid \chi_{R,i})$ for the unique $\chi_{R,i}$ true for $\vec{a}$. This is well-defined as the truth of a CPL formula in a structure only depends on the interpretation of the relation symbols occurring in the formula, and since for every structure and every choice of $\vec{a}$ there is a unique $\chi_{R,i}$ true for $\vec{a}$ by assumption.

Grounding to a Bayesian Network allows us to use all the inference methods developed for Bayesian networks. This includes conditioning on given data, and also lets us prescribe the interpretation of root predicates if desired. We will see examples of this in Subsection 3.1 below.

2.3 Functional lifted Bayesian networks

While LBN-CPL enable us to express more complex models with Type III conditions, they are intrinsically categorical: They do not allow the probability of $R(\vec{a})$ to vary as a *continuous* function of the Type-I-probabilities of first-order statements. Therefore we introduce a second formalism here, which is designed to do just that.

Definition 6. A *functional lifted Bayesian network (FLBN)* over a relational signature σ consists of the following:

1. A DAG G with node set σ.
2. For each $R \in \sigma$ a finite tuple $(\chi_{R,i}(\bar{x}, \bar{y}))_{i \leq n_R}$ of first-order $\mathrm{par}(R)$-formulas, where $|\bar{x}|$ is a sort-appropriate tuple of the correct length for R.
3. For each $R \in \sigma$ a continuous function $f_R : [0,1]^{n_R} \to [0,1]$.

The intuition behind FLBN is that the probability of $R(\overrightarrow{(a)})$ is given by the value of f applied to the tuple $(\|\chi_{R,i}\|_{\vec{y}})$. Note that if $\chi_{R,i}$ has no free variables beyond $\vec{x}$, then the Type I probability can only take the values 1 and 0, depending on whether $\chi_{R,i}$ is true or false.

Definition 7. Consider an FLBN $\mathfrak{G}$ and a finite domain D. Then the probability distribution induced by $\mathfrak{G}$ on the set of σ-structures with domain D is given by the following Bayesian network: The nodes are given by $R(\vec{a})$, where R is a relation symbol in σ and $\vec{a}$ is a tuple of elements of D of the right length and the right sorts for R. There is an edge between two nodes $R_1(\vec{a})$ and $R_2(\vec{b})$ if there is an edge between R_1 and R_2 in the DAG G underlying $\mathfrak{G}$. It remains to define a probability table for every node $R(\vec{a})$: Given a choice of values for $P(\vec{b})$ for all $P \in \mathrm{par}(R)$ and appropriate tuples $\vec{b}$ from D, the probability of $R(\vec{a})$ is set as $f_R((\|\chi_{R,i}\|_{\vec{y}})_{i \leq n_R})$.

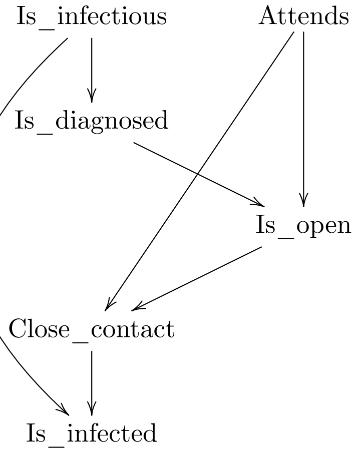
3 Discussion and Applications

In this section we will discuss how the Type III formalisms described here can be used to express dependencies beyond the existing Type II formalisms, which learning algorithms are supported and how they enable transfer learning across domains of different sizes.

3.1 Expressivity of lifted Bayesian networks for conditional probability logic

Continuing the running example of infectious disease dynamics, CPL allows the expression of various trigger conditions. CPL-LBN allow us to express the actions taken when those conditions are met. Overall, we can model each of the policy decisions summarised by Bisset et al. (2014, Table II).

Example. As an example combining several general features of statistical relational modelling with CPL, consider a policy where schools and workplaces are shut whenever there is either a diagnosed positive case in that school or workplace or when 0.1% of the population are diagnosed with the disease. Additionally, there is a higher chance of contact between two people if they attend the same open school or workplace. This can all be expressed in an LBN-CPL over a two-sorted signature as follows:



Assume that “Attends” is given by supplied data. Depending on whether this network is just one component in an iteration, where “Is_infectious” depends on the “Is_infected” of a previous time step, or stands alone, “Is_infectious” might be given by data. Alternatively, it might be stochastically modelled, with a certain fixed probability. Then the conditional probabilities of the other four relations can be expressed as follows: Is_diagnosed(x) has a given probability p_1 if Is_infectious(x) is true, and a (lower) probability p_2 otherwise.

Is_open(w) is deterministic, with probability 1 if

$$\exists_x(\text{Is\_diagnosed(x)} \wedge \text{Attends}(x, w)) \\ \vee \|\text{Is\_diagnosed}(x)\|_x \geq 0.01$$

is true, and with probability 0 otherwise.

Close_contact(x, y) models contact close enough to transmit the disease. It could be set at a probability p_3

if

$$\exists_w(\text{Is\_open}(w) \wedge \text{Attends}(x, w) \wedge \text{Attends}(y, w))$$

and at a (lower) probability p_4 otherwise. The values of p_3 and p_4 will be varied depending on the transmissivity of the disease and the social structure of the population.

Finally, $\text{Is\_Infected}(x)$ can now be seen as a deterministic dependency, with probability 1 if

$$\exists_y(\text{Close\_contact}(x, y) \wedge \text{Is\_Infected}(y))$$

and probability 0 otherwise.

We conclude this subsection by showing that unlike FLBN, every LBN-CPL can be expressed by one containing only probabilistic facts and deterministic rules expressed in CPL. More precisely:

Proposition 1. *For every LBN-CPL $\mathfrak{G}$ over a signature σ there is another LBN-CPL $\mathfrak{G}'$ in a signature $\sigma' \supseteq \sigma$ such that every relation R in σ' is either a root node or a child of only root nodes. In the latter case, all probabilities $\mu(R \mid \chi_{R,i})$ associated with R are either 0 or 1.*

Proof. The construction is similar to the expression of Bayesian logic programs by the distribution semantics (Riguzzi and Swift 2018, Section 4). For every relation symbol R of σ and every $\chi_{R,i}$, we add an additional relation symbol $P_{R,i}$ to σ of the same arity and sorts as R. This relation symbol is added as a root node and annotated with probability $\mu(R \mid \chi_{R,i})$. We also add an edge from this new node to R. Then we replace the $\mu(R \mid \chi_{R,i})$ with probability 1 and add an extra conjunction with 1 $P_{R,i}$ to every $\chi_{R,i}$. We then add an additional formula $\bigvee_i(\neg P_{R,i})$ and give it probability 0. As we iterate this construction through all nodes, starting at the roots, we can successively replace any mention of a non-root relation symbol with its definition, which is given by $\bigvee_i(\chi_{Q,i} \wedge P_{Q,i})$ and redraw the edges accordingly. Eventually, every relation symbol mentioned will be at the root of the DAG, as required. □

Note, however, that unlike in the case of PLP, the new relation $P(R_i)$ that we introduce has just the same arity and sorts as R, and is not in the scope of any quantifier.

3.2 Expressivity of Functional Lifted Bayesian Networks

While LBN-CPL are very good at modelling discrete trigger conditions, they are unsuitable for expressing general continuous dependencies on proportion. However, those dependencies are particularly important because they form the basis of the regression models from statistics. From this point of view, one could see FLBN as a general framework for relational regression models contingent on Type I probabilities as data. In particular, consider linear regression and logistic regression, two of the most commonly used regression functions.

A linear regression model with Type-I-probabilities as data corresponds to the following families of functions:

$$f(x_1, \ldots x_n) = \frac{w_1 x_1 + \ldots + w_n x_n}{n} + c \tag{1}$$

where w_1 to w_n and c are coefficients that have to be chosen in such a way that the image of $[0, 1]$ under f is contained in $[0, 1]$. These families of functions suffice e. g. to express the “arithmetic mean” combination function in RBN.

A logistic regression model can also be implemented; it specialises to DA-RLR. Its functions are of the form

$$f(x_1, \ldots x_n) = \text{sigmoid}(w_1 x_1 + \ldots + w_n x_n + c) \tag{2}$$

where “sigmoid” denotes the sigmoidal function

$$f(x) = \frac{\exp(x)}{\exp(x) + 1}$$

By recovering these existing frameworks and combination functions as special cases of FLBN, they are in scope of the rigorous analysis of asymptotic behaviour in Subsection 3.4.

3.3 Learning Type III formalisms

We will briefly consider the methods and algorithms that are available for learning LBN-CPL and FLBN. Let us consider LBN-CPL first. Here, parameter learning from interpretations amounts to learning the probabilities $\mu(R \mid \chi_{R,i})$ given one (or more) σ-structures. This case is closely aligned to parameter learning in other directed settings. In particular, in the case of a completely specified training model, the probabilities can be set to equal the frequencies encountered in the data. This can actually be expressed in terms of CPL itself: Set $\mu(R \mid \chi_{R,i})$ to be equal to the relative frequency $\|R(\vec{x})|\chi_{R,i}(\vec{x})\|_{\vec{x}}$. In the case of missing data, we can adapt the expectation-maximisation algorithm for learning parameters in ordinary Bayesian networks to CPL. Since we can view the LBN-CPL as a deterministic model over probabilistic facts (Proposition 1), this can be done by tying parameters known to be the same from the relational structure of the model and the formulas $\chi_{R,i}$ in just the same way as in probabilistic inductive logic programming (outlined for instance by de Raedt (2008, Section 8.5)). This native support for parameter learning makes it particularly attractive for modelling tasks such as the ones outlined above. In the context of infectious disease dynamics for instance, the parameters have epidemiological meanings and their estimation from past data is an important part of informed decision making. While structure learning could in principle be accomplished along the lines of the SLIPCOVER beam search algorithm (Bellodi and Riguzzi 2015), we anticipate that the significantly larger theory space resulting from the added probability quantifiers would make that difficult. Indeed, as far as we are aware even the problem of plain inductive logic programming in CPL without a Type III extension is yet to be addressed.

When posing the problem of parameter learning for FLBN, we first have to clarify which parameters we are learning in the first place. So rather than just considering functions $f : [0,1]^n \to [0,1]$ we consider a parametrised family of such functions, $f : K \times [0,1]^n \to [0,1]$, where K is the closure of a connected open subset of $\mathbb{R}_\infty^m$ from which the parameters are taken (Here $\mathbb{R}_\infty$ stands for $\mathbb{R} \cup \{-\infty, \infty\}$). For instance, in the cases of linear and logistic regression, Equations 1 and 2 define functions taking $m = n + 1$ parameters, $w_1, \ldots, w_n$ and c. In the logistic case, $K = \mathbb{R}^m$, while in the linear case, the parameter space is constrained by the function mapping to $[0,1]$. Jaeger (2007) presents a general approach for learning the parameters of aggregation functions using gradient descent whenever the functions are differentiable in the parameters. Clearly both the linear and the logistic regression examples are differentiable in the parameters. Functional gradient boosting has been successfully applied to the structure learning of relational logistic regression by Ramanan et al. (2018), and it seems very promising to evaluate this approach with other classes of regression functions expressible by FLBN. We believe structure learning in FLBN to be much more feasible than for lifted Bayesian networks for CPL. Firstly, there is a large bank of work on regression learning in the statistical literature on which relational versions could be based, and secondly, the scale of the task can be reduced systematically by partly specifying families of functions (recovering e. g. structure learning in relational logistic regression or linear regression as special cases).

3.4 Asymptotic Analysis of the Extrapolation Behaviour

Utilising Koponen's results (2020), we can give a full analysis of the asymptotic behaviour of the two formalisms presented for Type III logic presented here. The setting is as follows: On every finite domain D, a lifted Bayesian network $\mathfrak{G}$ over a signature σ induces a probability distribution $\mathbb{P}_{\mathfrak{G},D}$ on the set Ω_D of σ-structures with domain D. The first thing to note is that the names of the elements of D do not matter; all relevant information lies in the cardinalities of the sorts of D. Therefore we will assume from now on that our domain sorts consist of initial segments of the natural numbers, and we will write $\mathbb{P}_{\mathfrak{G},\vec{n}}$ for the probability distribution on the sorts with $\vec{n}$ elements. In an asymptotic analysis, we are interested in the limit of these probability distributions as the domain size of D tend to infinity (If σ is multi-sorted, we require the size of every sort to tend to infinity). A technical difficulty here is that strictly speaking, the probability distributions are defined on different sets Ω_n, so it is unclear on which measure space a limit would even be defined. To be precise, we consider the measure space Ω_∞ given by all σ-structures with domain sorts $\mathbb{N}$. It is endowed with the σ-algebra generated by *generating sets* of the following form: "Those σ-structures $\mathfrak{X}$ such that the σ-substructure of $\mathfrak{X}$ with domain $a_1, \ldots a_m$ is given by $\mathfrak{Y}$" for a tuple of domain elements $a_1, \ldots a_m$ and a σ-structure $\mathfrak{Y}$. We can identify this set with $\mathfrak{Y}$. This suffices to give a probability to any query about finite domains. On such a generating set, all but finitely many $\mathbb{P}_{\mathfrak{G},\vec{n}}$ are defined; indeed, $\mathbb{P}_{\mathfrak{G},\vec{n}}$ gives a probability to any structure with domain $a_1, \ldots, a_m$ as long as every a_i is bounded by the entry of $\vec{n}$ corresponding to its sort. Furthermore, the probability of the generating sets completely determine the probability distribution on the measure space itself. Thus, we can make the following definitions:

Definition 8. Two formulas of CPL φ and ψ are *asymptotically equivalent* over a lifted Bayesian network if for any sequence D_k of domains which is monotone and unbounded in the cardinality of every sort $\lim_{k\to\infty} \mathbb{P}_{\mathfrak{G},D_k}(\forall_{\vec{x}} \varphi(\vec{x}) \leftrightarrow \psi(\vec{x})) = 1$

A probability distribution $\mathbb{P}_\infty$ on Ω_∞ is the *asymptotic limit* of a lifted Bayesian network $\mathfrak{G}$ if for any sequence D_k of domains which is monotone and unbounded in the cardinality of every sort and any generating set A the limit of $\mathbb{P}_{\mathfrak{G},\vec{n}}(A)$ equals $\mathbb{P}_\infty(A)$. Two lifted Bayesian networks $\mathfrak{G}$ and $\mathfrak{G}'$ are *asymptotically equivalent* if they share the same asymptotic limit $\mathbb{P}_\infty$.

The discussion above does not imply, of course, that a given lifted Bayesian network actually has an asymptotic limit in that sense. However, there is a class of lifted Bayesian networks where the asymptotic limit is clear: those that define *projective families of distributions* in the sense of Jaeger and Schulte (2018; 2020).

Definition 9. A family of probability distributions $(\mathbb{P}_{\vec{n}})$ on $\Omega_{\vec{n}}$ is *projective* if for every generating set A of Ω_∞ the sequence $(\mathbb{P}_{\vec{n}}(A))$ is constant whenever it is defined.

Since clearly every constant sequence converges, every lifted Bayesian network inducing a projective family of distributions has an asymptotic limit. Furthermore, if two families of distributions are asymptotically equivalent and both projective, they must be equal. This leads us to the following observation:

Proposition 2. *If every formula $\chi_{R,i}$ in an LBN-CPL $\mathfrak{G}$ is a quantifier-free first-order formula (a* quantifier-free lifted Bayesian network*), then $\mathfrak{G}$ induces a projective family of distributions. In particular, $\mathfrak{G}$ has an asymptotic limit.*

Proof. Since the construction in the proof of Proposition 1 does not add any new quantifiers, we can assume without loss of generality that $\mathfrak{G}$ is of the form described there. Then the probability of any choice of truth values for relations applied to

$a_1, \dots, a_m$ depends only on the probabilities of their definitions $\chi_{R,i}$, which only mention relations at root nodes. As the $\chi_{R,i}$ are Boolean combinations of atomic formulas, which are independent since they are all at root nodes in the DAG, their probabilities can be obtained directly from the probabilities of the atomic formulas, which in turn are specified at the root node itself independently of D. ☐

With these preparations out of the way, we can formulate our main result on the asymptotic behaviour of FLBN.

Theorem 1. *Let $\mathfrak{G}$ be an FLBN. Then $\mathfrak{G}$ is asymptotically equivalent to a quantifier-free lifted Bayesian Network $\mathfrak{G}'$.*

The proof builds substantially on Koponen's analysis of LBN-CPL (2020). Koponen introduces the notions of a *critical* number and a *critical* formula. We will not require their technical definitions here; it suffices for us to know that the definitions depend (only) on $\mathfrak{G}$, that every first-order formula is non-critical and that for all conditional probability formulas $\varphi, \psi, \theta, \tau$ and tuples of variables $\vec{y}$ for all but finitely many r the formulas $r + \|\varphi|\psi\|_{\vec{y}} \geq \|\theta|\tau\|_{\vec{y}}$ and $\|\varphi|\psi\|_{\vec{y}} \geq \|\theta|\tau\|_{\vec{y}} + r$ are non-critical. Note that the asymptotic characterisation of FLBN does not depend on any assumption of non-criticality or similar in its statement. Koponen's (2020) main results are then (Theorems 3.14 and 3.16 respectively):

Theorem 2. *Over any given LBN-CPL $\mathfrak{G}$, every non-critical conditional probability formula φ is asymptotically equivalent to a quantifier-free formula ψ.*

On the level of networks, Koponen obtains a similar asymptotic convergence result:

Theorem 3. *Any LBN-CPL $\mathfrak{G}$ all of whose partition formulas $\chi_{R,i}$ are non-critical is asymptotically equivalent to a quantifier-free lifted Bayesian network.*

While Koponen (2020) gives a general proof, the construction is straightforward for lifted Bayesian networks in the form of Proposition 1:

Proof. Let $\mathfrak{G}$ be a lifted Bayesian network in the form of Proposition 1. Then every $\chi_{R,i}$ is asymptotically equivalent to a quantifier-free first-order formula over the distribution defined by the probabilistic facts, which give the required quantifier-free representation. ☐

Now we are ready to sketch a proof of Theorem 1.

Proof. Let R be a relation symbol from σ and let σ' be the signature of ancestors of R. Then the analysis of Theorem 2 can be used to find a partition of quantifier-free formulas $\{\chi'_{R,i}\} \in CPL(\sigma')$ such that for any unbounded increasing sequence of σ' structures that satisfy $\chi'_{r,i}$, the tuple $(\|\chi_{R,i}\|_{\vec{y}})_{i \leq n_R}$ converges to a particular $(c_{R,i})_{i \leq n_R}$. Since f is continuous, this implies that the limit of $f((\|\chi_{R,i}\|_{\vec{y}})_{i \leq n_R})$ is $f(c_{\vec{R},i})$. Therefore the network with the partition $\{\chi'_{R,i}\}$ and the probabilities $\mu(R|\chi'_{R,i}) \coloneqq f(c_{\vec{R},i})$ is asymptotically equivalent.

We close by remarking how our asymptotic results complement those of Jaeger (1998). There, Jaeger shows that the probability distributions of relational Bayesian networks with *exponentially convergent* combination functions lead to asymptotically convergent probability distributions. However, the exponentially convergent combination are essentially those that given a certain type of input sequence increasing in length converge to a fixed value, regardless of the precise sequence received. The classical combination functions 'noisy-or' and 'maximum' are paradigmal for this behaviour. The central idea of of our work here is that the functions converge to a value that depends explicitly on the means of the sequences received, and therefore they are clearly distinguished from Jaeger's exponentially convergent combination functions. ☐

3.5 Examples

We illustrate the analysis of the last subsection with a sequence of simple examples that serve to highlight the main aspects. Consider the situation of Example 2: The signature σ has two unary relation symbols Q and R, and the underlying DAG G is $Q \longrightarrow R$. We model a relationship between $R(x)$ and those $y \in D$ that satisfy $Q(y)$. In Example 2, we have seen an RLR approach to this problem. Here, the asymptotic behaviour is well-known: as domain size increases, the expected number of $a \in D$ that satisfy $R(a)$ also does. By the law of large numbers, this increase is almost surely linear with domain size. Therefore the probability of $Q(y)$ will limit to 0 if and only if $w < 0$, and limit to 1 if and only if $w > 0$. A similar analysis holds if we consider a noisy-or combination, or the model of a probabilistic logic program.

So now consider modelling such a dependency with an FLBN.

Example 3. Let $\mathfrak{G}$ be an FLBN on G with a formula $Q(y)$ and a function $f : [0,1] \to [0,1]$. Assume further that $\mu(Q) \in [0,1]$. Then just as in Example 4, $\|Q(y)\|_y$ converges to $\mu(Q)$ almost surely as domain size increases. Since f is continuous, this implies that $f(\|Q(y)\|_y)$ converges to $f(\mu(Q))$ almost surely. So the asymptotically equivalent quantifier-free network will also have "true" as its formula for R and then $f(\mu(Q))$ as $\mu(R|\text{true})$.

If we assume the dependency to be discrete with a known cut-off point r in the relative frequency of R, we could consider using LBN-CPL to model it:

Example 4. Let $\mathfrak{G}$ be an LBN-CPL on G with a probability $\mu(Q) \in [0,1]$ for Q and two conditional probability formulas characterising R: $\chi_{R,1} \coloneqq \|Q(y)\|_y \geq r$ and $\chi_{R,2} \coloneqq \neg(\|Q(y)\|_y \geq r)$, where $r \in [0,1]$. We furthermore choose $\mu(R|\chi_{R,1}) \in [0,1]$

and $\mu(R|\chi_{R,i}) \in [0,1]$. It turns out that the formulas $\chi_{R,i}$ are non-critical only if $r \neq \mu(Q)$. The asymptotic analysis here proceeds as follows: By the law of large numbers, $\|Q(y)\|_y$ will be almost surely arbitrarily close to $\mu(Q)$ as domain size increases, and therefore almost surely $\chi_{R,1}$ will be true if and only $r < \mu(Q)$. Thus the asymptotically equivalent quantifier-free Bayesian network will simply have the quantifier-free formula "true" as χ_R and $\mu(R)$ will be $\mu(R|\chi_{R,1})$ if $r < \mu(Q)$ and $\mu(R|\chi_{R,2})$ if $r > \mu(Q)$. Note that we cannot make any statement about the critical case $r = \mu(Q)$. Alternatively, consider the representation in the form of Proposition 1, which is as follows: The signature σ' has two additional unary predicates P_1 and P_2, with $\mu(P_i) := \mu(R|\chi_{R,i})$. The definition of $R(x)$ is then given by $R(x) := (\chi_{R,1}(x) \wedge P_1(x)) \vee (\chi_{R,2}(x) \wedge P_2(x))$. Asymptotically, $\chi_{r,1}(x)$ and $\chi_{r,1}(x)$ behave just as in the original representation, so the asymptotic representation will be $R(x) := P_1(x)$ if $r < \mu(Q)$ and $R(x) := P_2(x)$ if $r > \mu(Q)$.

3.6 Transfer learning across domain sizes

As projective families of distributions, quantifier-free lifted Bayesian networks have very desirable properties for learning across domain sizes. More precisely, for the family of distributions induced by any quantifier-free Bayesian network and any structure $\mathfrak{X}$ with $m < n$ elements, $\mathbb{P}_n(\mathfrak{X}) = \mathbb{P}_m(\mathfrak{X})$.

Consider a parametric family of distributions G_θ which are asymptotically equivalent to a parametric projective family of distributions G'_θ. Consider the problem of learning the parameters from interpretations on a structure $\mathfrak{X}$ of large domain size n. Then we could proceed as follows: sample substructures of domain size $m < n$, where m is larger than the highest arity in σ and the arity of the queries we are typically interested in. Find the parameters of G'_θ that maximise the sum of the log-likelihoods of the samples of size m. Now consider G_θ. By the asymptotic convergence results, if n is sufficiently large, these parameters maximise the likelihood of obtaining the substructures of size m sampled from $\mathfrak{X}$ using G_θ, including realisations the typical queries. This procedure is related to that of Kuzelka et al. (2018). However, while they consider samples of size m as training data, they still learn with respect to a fixed sample size n, which is necessary because their setting has no assumption of (asymptotic) projectivity.

Estimating the parameters on the parametric families G'_θ only works if G'_θ truly depends on the parameters, and does so in a regular way; for the learning algorithms suggested here that means that $G'_\theta(\mathfrak{X})$ should be differentiable in the parameters.

Let us evaluate these criteria for the asymptotic approximations in Examples 2 - 4 above:

In the case of Example 2, G'_θ does not depend on w beyond its sign. Therefore, this asymptotic representation is unsuitable for learning the parameters.

In the lifted functional Bayesian network model of Example 3, the parametric family is defined by $f_\theta(\mu_Q)$. If f is linear or logistic, for instance, then $f_\theta(\mu_Q)$ will depend appropriately on the parameters for any fixed $\mu_Q \in (0,1)$. Note, that while $f(\mu_Q)$ will vary with every parameter individually, it will take its maximum-likelihood value (which happens to coincide with the true frequency of $R(x)$ in the domain) on an infinite subspace of tuples of parameters. This is not unique to the projective approximation, however, but is a well-known phenomenon when learning the parameters for a relational logistic regression from a single interpretation (see (Kazemi et al. 2014) and (Poole et al. 2014) for an in-depth discussion). This can be overcome by learning from several large domains, where Q has different frequencies.

In the conditional probability modelling of Example 4, we could start by estimating μ_Q from data on the values of Q. If $\mu_Q \neq r$, we can then proceed to learn one of the parameters using the asymptotic limit as outlined above. In this parameter, the dependence is clearly linear and therefore satisfies all of the conditions. However, the other parameter does not occur in the projective limit and therefore cannot be estimated in this way. To estimate that parameter also, we would also need more training domains, including those where the relative frequency of Q is above the threshold

In this way, Type III formalisms allow us to leverage the power of projective families of distributions for transfer learning while retaining much more expressive modelling capabilities. While adding either functional or discrete dependencies on the Type I probabilities present in a domain allow us to express rich connections between different domain elements, quantifier-free lifted Bayesian networks themselves do not allow any dependence on the global structure of the model. This is also quite typical of projective families of distributions that can be expressed in statistical relational AI, since any projective LBN-CPL for instance is expressible by a quantifier-free one (as two asymptotically equivalent projective families of distributions are completely equivalent).

3.7 Conclusion

LBN-CPL and FLBN introduce relative frequencies into statistical relational artificial intelligence, making Halpern Type III probabilities available to this field. By supporting discrete and continuous dependencies on relative frequencies, they can express the complex relationships that are required to model application domains such as infectious disease dynamics. LBN-CPL and FLBN also advance statistical relational learning from large interpretations by supporting learning from randomly sampled subdomains. This is underpinned by a rigorous analysis of their asymptotic behaviour. Furthermore, the transparent relationship to Bayesian networks via their grounded semantics allows the application of well-developed learning and inference approaches.

Acknowledgements

We would like to thank Vera Koponen for several insightful conversations, and François Bry and Kailin Sun for their very helpful comments on the manuscript.

References

Bellodi, E., and Riguzzi, F. 2015. Structure learning of probabilistic logic programs by searching the clause space. *Theory Pract. Log. Program.* 15(2):169–212.

Bisset, K. R.; Chen, J.; Deodhar, S.; Feng, X.; Ma, Y.; and Marathe, M. V. 2014. Indemics: An interactive high-performance computing framework for data-intensive epidemic modeling. *ACM Trans. Model. Comput. Simul.* 24(1):4:1–4:32.

Carnap, R. 1950. *Logical Foundations of Probability.* University of Chicago Press.

Dalvi, N. N., and Suciu, D. 2012. The dichotomy of probabilistic inference for unions of conjunctive queries. *J. ACM* 59(6):30:1–30:87.

Doan, A.; Niu, F.; Ré, C.; Shavlik, J.; and Zhang, C. 2011. *User manual of Tuffy 0.3.* Stanford University.

Ebbinghaus, H.; Flum, J.; and Thomas, W. 1994. *Mathematical logic (2. ed.).* Undergraduate texts in mathematics. Springer.

Halpern, J. Y. 1990. An analysis of first-order logics of probability. *Artif. Intell.* 46(3):311–350.

Jaeger, M., and Schulte, O. 2018. Inference, learning, and population size: Projectivity for SRL models. In *Eighth International Workshop on Statistical Relational AI (StarAI).* arXiv:1807.00564.

Jaeger, M., and Schulte, O. 2020. A complete characterization of projectivity for statistical relational models. In Bessiere, C., ed., *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, 4283–4290. ijcai.org.

Jaeger, M. 1998. Convergence results for relational bayesian networks. In *Thirteenth Annual IEEE Symposium on Logic in Computer Science, Indianapolis, Indiana, USA, June 21-24, 1998*, 44–55. IEEE Computer Society.

Jaeger, M. 2002. Relational bayesian networks: A survey. *Electron. Trans. Artif. Intell.* 6:1–21.

Jaeger, M. 2007. Parameter learning for relational bayesian networks. In Ghahramani, Z., ed., *Machine Learning, Proceedings of the Twenty-Fourth International Conference (ICML 2007), Corvallis, Oregon, USA, June 20-24, 2007*, volume 227 of *ACM International Conference Proceeding Series*, 369–376. ACM.

Kazemi, S. M.; Buchman, D.; Kersting, K.; Natarajan, S.; and Poole, D. 2014. Relational logistic regression. In Baral, C.; Giacomo, G. D.; and Eiter, T., eds., *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014.* AAAI Press.

Koponen, V. 2020. Conditional probability logic, lifted bayesian networks, and almost sure quantifier elimination. *Theor. Comput. Sci.* 848:1–27.

Kuzelka, O.; Wang, Y.; Davis, J.; and Schockaert, S. 2018. Relational marginal problems: Theory and estimation. In McIlraith, S. A., and Weinberger, K. Q., eds., *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, 6384–6391. AAAI Press.

Muggleton, S., and Chen, J. 2008. A behavioral comparison of some probabilistic logic models. In Raedt, L. D.; Frasconi, P.; Kersting, K.; and Muggleton, S., eds., *Probabilistic Inductive Logic Programming - Theory and Applications*, volume 4911 of *Lecture Notes in Computer Science*. Springer. 305–324.

Poole, D.; Buchman, D.; Kazemi, S. M.; Kersting, K.; and Natarajan, S. 2014. Population size extrapolation in relational probabilistic modelling. In Straccia, U., and Calì, A., eds., *Scalable Uncertainty Management - 8th International Conference, SUM 2014, Oxford, UK, September 15-17, 2014. Proceedings*, volume 8720 of *Lecture Notes in Computer Science*, 292–305. Springer.

Raedt, L. D. 2008. *Logical and relational learning.* Cognitive Technologies. Springer.

Ramanan, N.; Kunapuli, G.; Khot, T.; Fatemi, B.; Kazemi, S. M.; Poole, D.; Kersting, K.; and Natarajan, S. 2018. Structure learning for relational logistic regression: An ensemble approach. In Thielscher, M.; Toni, F.; and Wolter, F., eds., *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018*, 661–662. AAAI Press.

Richardson, M., and Domingos, P. M. 2006. Markov logic networks. *Mach. Learn.* 62(1-2):107–136.

Riguzzi, F., and Swift, T. 2018. A survey of probabilistic logic programming. In Kifer, M., and Liu, Y. A., eds., *Declarative Logic Programming: Theory, Systems, and Applications.* ACM / Morgan & Claypool. 185–228.

Schulte, O.; Khosravi, H.; Kirkpatrick, A. E.; Gao, T.; and Zhu, Y. 2014. Modelling relational statistics with bayes nets. *Mach. Learn.* 94(1):105–125.

Schulte, O. 2012. Challenge paper: Marginal probabilities for instances and classes. Presented at the ICML-SRL Workshop on Statistical Relational Learning.

Weitkämper, F. 2021. Scaling the weight parameters in markov logic networks and relational logistic regression models. arXiv:2103.12140.