Markov Logic in Infinite Domains

Parag Singla **Pedro Domingos**
Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195-2350, U.S.A.
{*parag, pedrod*}*@cs.washington.edu*

Abstract

Combining first-order logic and probability has long been a goal of AI. Markov logic (Richardson & Domingos, 2006) accomplishes this by attaching weights to first-order formulas and viewing them as templates for features of Markov networks. Unfortunately, it does not have the full power of first-order logic, because it is only defined for finite domains. This paper extends Markov logic to infinite domains, by casting it in the framework of Gibbs measures (Georgii, 1988). We show that a Markov logic network (MLN) admits a Gibbs measure as long as each ground atom has a finite number of neighbors. Many interesting cases fall in this category. We also show that an MLN admits a unique measure if the weights of its non-unit clauses are small enough. We then examine the structure of the set of consistent measures in the non-unique case. Many important phenomena, including systems with phase transitions, are represented by MLNs with non-unique measures. We relate the problem of satisfiability in first-order logic to the properties of MLN measures, and discuss how Markov logic relates to previous infinite models.

1 Introduction

Most AI problems are characterized by both uncertainty and complex structure, in the form of multiple interacting objects and relations. Handling both requires combining the capabilities of probabilistic models and first-order logic. Attempts to achieve this have a long history, and have gathered steam in recent years. Within AI, Nilsson (1986) is an early example. Bacchus (1990), Halpern (1990) and coworkers (e.g., Bacchus *et al.* (1996)) produced a substantial body of relevant theoretical work. Around the same time, several authors began using logic programs to compactly specify complex Bayesian networks, an approach known as knowledge-based model construction (Wellman et al., 1992). More recently, many combinations of (subsets of) first-order logic and probability have been proposed in the burgeoning field of statistical relational learning (Dietterich et al., 2003), including probabilistic relational models (Friedman et al., 1999), stochastic logic programs (Muggleton, 1996), Bayesian logic programs (Kersting & De Raedt, 2001), and others.

One of the most powerful representations to date is Markov logic (Richardson & Domingos, 2006). Markov logic is a simple combination of Markov networks and first-order logic: each first-order formula has an associated weight, and each grounding of a formula becomes a feature in a Markov network, with the corresponding weight. The use of Markov networks instead of Bayesian networks obviates the difficult problem of avoiding cycles in all possible groundings of a relational model (Taskar et al., 2002). The use of first-order logic instead of more limited representations (e.g., description logics, Horn clauses) makes it possible to compactly represent a broader range of dependencies. For example, a dependency between relations like "Friends of friends are (usually) friends" cannot be specified compactly in (say) probabilistic relational models, but in Markov logic it suffices to write down the corresponding formula and weight. Markov logic has been successfully applied in a variety of domains (Domingos et al., 2006), and open source software with implementations of state-of-the-art inference and learning algorithms for it is available (Kok et al., 2006).

One limitation of Markov logic is that it is only defined for finite domains. Thus the question of how far it is possible to combine the full power of first-order logic and graphical models remains open. This paper takes a further step in this direction by extending Markov logic to infinite domains. Our treatment is based on the theory of Gibbs measures (Georgii, 1988). Gibbs measures are infinite-dimensional extensions of Markov networks, and have been studied extensively by statistical physicists and mathematical statisticians, due to their importance in modeling systems with phase transitions. We begin with some necessary background on first-order logic and Gibbs measures. We then define MLNs over infinite domains, state sufficient conditions for the existence and uniqueness of a probability measure consistent with a given MLN, and examine the important case of MLNs with non-unique measures. Next, we establish a correspondence between the problem of satisfiability in logic and the existence of MLN measures with certain properties. We conclude with a discussion of the relationship between infinite MLNs and previous infinite relational models.

2 Background

2.1 First-Order Logic

A *first-order knowledge base* is a set of sentences or formulas in first-order logic (Genesereth & Nilsson, 1987). Formulas are constructed using four types of symbols: constants, variables, functions, and predicates. Constant symbols represent objects in the domain of discourse (e.g., people: `Anna`, `Bob`, `Chris`, etc.). Variable symbols range over the objects in the domain (or a subset of it, in which case they are *typed*). Function symbols (e.g., `MotherOf`) represent mappings from tuples of objects to objects. Predicate symbols represent relations among objects (e.g., `Friends`) or attributes of objects (e.g., `Smokes`). A *term* is any expression representing an object. It can be a constant, a variable, or a function applied to a tuple of terms. For example, `Anna`, `x`, and $\mathtt{GreatestCommonDivisor}(\mathtt{x}, \mathtt{y})$ are terms. An *atomic formula* or *atom* is a predicate symbol applied to a tuple of terms (e.g., $\mathtt{Friends}(\mathtt{x}, \mathtt{MotherOf}(\mathtt{Anna}))$). A *ground term* is a term containing no variables. A *ground atom* or *ground predicate* is an atomic formula all of whose arguments are ground terms. Formulas are recursively constructed from atomic formulas using logical connectives and quantifiers. A *positive literal* is an atomic formula; a *negative literal* is a negated atomic formula. A *clause* is a disjunction of literals. Every first-order formula can be converted into an equivalent formula in *prenex conjunctive normal form*, $Qx_1 \ldots Qx_n C(x_1, \ldots, x_n)$, where each Q is a quantifier, the x_i are the quantified variables, and $C(\ldots)$ is a conjunction of clauses.

The *Herbrand universe* $\mathbf{U}(\mathbf{C})$ of a set of clauses $\mathbf{C}$ is the set of all ground terms constructible from the function and constant symbols in $\mathbf{C}$ (or, if $\mathbf{C}$ contains no constants, some arbitrary constant, e.g., `A`). If $\mathbf{C}$ contains function symbols, $\mathbf{U}(\mathbf{C})$ is infinite. (For example, if $\mathbf{C}$ contains solely the function `f` and no constants, $\mathbf{U}(\mathbf{C}) = \{\mathtt{f}(\mathtt{A}), \mathtt{f}(\mathtt{f}(\mathtt{A})), \mathtt{f}(\mathtt{f}(\mathtt{f}(\mathtt{A}))), \ldots\}$.) Some authors define the *Herbrand base* $\mathbf{B}(\mathbf{C})$ of $\mathbf{C}$ as the set of all ground atoms constructible from the predicate symbols in $\mathbf{C}$ and the terms in $\mathbf{U}(\mathbf{C})$. Others define it as the set of all ground clauses constructible from the clauses in $\mathbf{C}$ and the terms in $\mathbf{U}(\mathbf{C})$. For convenience, in this paper we will define it as the union of the two, and talk about the *atoms in* $\mathbf{B}(\mathbf{C})$ and *clauses in* $\mathbf{B}(\mathbf{C})$ as needed.

An *interpretation* is a mapping between the constant, predicate and function symbols in the language and the objects, functions and relations in the domain. In a *Herbrand interpretation* there is a one-to-one mapping between ground terms and objects (i.e., every object is represented by some ground term, and no two ground terms correspond to the same object). A *model* or *possible world* specifies which relations hold true in the domain. Together with an interpretation, it assigns a truth value to every atomic formula, and thus to every formula in the knowledge base.

2.2 Gibbs Measures

Gibbs measures are infinite-dimensional generalizations of Gibbs distributions. A Gibbs distribution, also known as a log-linear model or exponential model, and equivalent under mild conditions to a Markov network or Markov random field, assigns to a state $\mathbf{x}$ the probability

$$P(\mathbf{X}=\mathbf{x}) = \frac{1}{Z} \sum_i w_i f_i(\mathbf{x}) \qquad (1)$$

where w_i is any real number, f_i is an arbitrary function or *feature* of $\mathbf{x}$, and Z is a normalization constant. In this paper we will be concerned exclusively with Boolean states and functions (i.e., states are binary vectors, corresponding to possible worlds, and functions are logical formulas). Markov logic can be viewed as the use of first-order logic to compactly specify families of these functions (Richardson & Domingos, 2006). Thus, a natural way to generalize it to infinite domains is to use the existing theory of Gibbs measures (Georgii, 1988). Although Gibbs measures were primarily developed to model regular lattices (e.g., ferromagnetic materials, gas/liquid phases, etc.), the theory is quite general, and applies equally well to the richer structures definable using Markov logic.

One problem with defining probability distributions over infinite domains is that the probability of most or all worlds will be zero. Measure theory allows us to overcome this problem by instead assigning probabilities to sets of worlds (Billingsley, 1995). Let Ω denote the set of all possible worlds, and $\mathcal{E}$ denote a set of subsets of Ω. $\mathcal{E}$ must be a σ-algebra, i.e., it must be non-empty and closed under complements and countable unions. A function $\mu : \mathcal{E} \to \mathbb{R}$ is said to be a *probability measure* over $(\Omega, \mathcal{E})$ if $\mu(E) \geq 0$ for every $E \in \mathcal{E}$, $\mu(\Omega) = 1$, and $\mu(\bigcup E_i) = \sum \mu(E_i)$, where the union is taken over any countable collection of disjoint elements of $\mathcal{E}$.

A related difficulty is that in infinite domains the sum in Equation 1 may not exist. However, the distribution of any finite subset of the state variables conditioned on its complement is still well defined. We can thus define the infinite distribution indirectly by means of an infinite collection of finite conditional distributions. This is the basic idea in Gibbs measures.

Let us introduce some notation which will be used throughout the paper. Consider a countable set of variables $\mathbf{S} = \{X_1, X_2, \ldots\}$, where each X_i takes values in $\{0, 1\}$. Let $\mathbf{X}$ be a finite set of variables in $\mathbf{S}$, and $\mathbf{S_X} = \mathbf{S} \setminus \mathbf{X}$. A possible world $\omega \in \Omega$ is an assignment to all the variables in $\mathbf{S}$. Let $\omega_\mathbf{X}$ denote the assignment to the variables in $\mathbf{X}$ under ω, and ω_{X_i} the assignment to X_i. Let $\mathcal{X}$ denote the set of all finite subsets of $\mathbf{S}$. A *basic event* $\mathbf{X} = \mathbf{x}$ is an assignment of values to a finite subset of variables $\mathbf{X} \in \mathcal{X}$, and denotes the set of possible worlds $\omega \in \Omega$ such that $w_\mathbf{X} = \mathbf{x}$. Let $\mathbf{E}$ be the set of all basic events, and let $\mathcal{E}$ be the set obtained by taking complements and countable unions of sets in $\mathbf{E}$. By definition, $\mathcal{E}$ is a σ-algebra. An element E of $\mathcal{E}$ is called an *event*, and $\mathcal{E}$ is the *event space*. The following treatment is adapted from Georgii (1988).

Definition 1. *An* interaction potential *(or simply a* potential*) is a family $\Phi = (\Phi_\mathbf{V})_{\mathbf{V}\in\mathcal{X}}$ of functions $\Phi_\mathbf{V} : \mathbf{V} \to \mathbb{R}$ such that, for all $\mathbf{X} \in \mathcal{X}$ and $\omega \in \Omega$, the summation*

$$H^\Phi_\mathbf{X}(\omega) = \sum_{\mathbf{V}\in\mathcal{X}, \mathbf{V}\cap\mathbf{X}\neq\emptyset} \Phi_\mathbf{V}(\omega_\mathbf{V}) \tag{2}$$

is finite. $H^\Phi_\mathbf{X}$ is called the Hamiltonian in $\mathbf{X}$ for Φ.

Intuitively, the Hamiltonian $H^\Phi_\mathbf{X}$ includes a contribution from all the potentials $\Phi_\mathbf{V}$ which share at least one variable with the set $\mathbf{X}$. Given an interaction potential Φ and a subset of variables $\mathbf{X}$, we define the conditional distribution $\gamma^\Phi_\mathbf{X}(\mathbf{X}|\mathbf{S_X})$ as[1]

$$\gamma^\Phi_\mathbf{X}(\mathbf{X} = \mathbf{x}|\mathbf{S_X} = \mathbf{y}) = \frac{\exp(H^\Phi_\mathbf{X}(\mathbf{x}, \mathbf{y}))}{\sum_{\mathbf{x}\in\mathrm{Dom}(\mathbf{X})} \exp(H^\Phi_\mathbf{X}(\mathbf{x}, \mathbf{y}))} \tag{3}$$

where the denominator is called the *partition function* in $\mathbf{X}$ for Φ and denoted by $Z^\Phi_\mathbf{X}$, and $\mathrm{Dom}(\mathbf{X})$ is the domain of $\mathbf{X}$. Equation 3 can be easily extended to arbitrary events $E \in \mathcal{E}$ by defining $\gamma^\Phi_\mathbf{X}(E|\mathbf{S_X})$ to be non-zero only when E is consistent with the assignment in $\mathbf{S_X}$. Details are skipped here to keep the discussion simple, and can be found in Georgii (1988). The family of conditional distributions $\gamma^\Phi = (\gamma^\Phi_\mathbf{X})_{\mathbf{X}\in\mathcal{X}}$ as defined above is called a *Gibbsian specification*.[2]

Given a measure μ over $(\Omega, \mathcal{E})$ and conditional probabilities $\gamma^\Phi_\mathbf{X}(E|\mathbf{S_X})$, let the composition $\mu\gamma^\Phi_\mathbf{X}$ be defined as

$$\mu\gamma^\Phi_\mathbf{X}(E) = \int_{\mathrm{Dom}(\mathbf{S_X})} \gamma^\Phi_\mathbf{X}(E|\mathbf{S_X})\, \partial\mu \tag{4}$$

$\mu\gamma^\Phi_\mathbf{X}(E)$ is the probability of event E according to the conditional probabilities $\gamma^\Phi_\mathbf{X}(E|\mathbf{S_X})$ and the measure μ on $\mathbf{S_X}$. We are now ready to define Gibbs measure.

[1] For physical reasons, this equation is usually written with a negative sign in the exponent, i.e., $\exp[-H^\Phi_\mathbf{X}(\omega)]$. Since this is not relevant in Markov logic and does not affect any of the results, we omit it.

[2] Georgii (1988) defines Gibbsian specifications in terms of underlying independent specifications. For simplicity, we assume these to be equidistributions and omit them throughout this paper.

Definition 2. *Let γ^Φ be a Gibbsian specification. Let μ be a probability measure over the measurable space $(\Omega, \mathcal{E})$ such that, for every $\mathbf{X} \in \mathcal{X}$ and $E \in \mathcal{E}$, $\mu(E) = \mu\gamma^\Phi_\mathbf{X}(E)$. Then the specification γ^Φ is said to admit the* Gibbs measure *μ. Further, $\mathcal{G}(\gamma^\Phi)$ denotes the set of all such measures.*

In other words, a Gibbs measure is consistent with a Gibbsian specification if its event probabilities agree with those obtained from the specification. Given a Gibbsian specification, we can ask whether there exists a Gibbs measure consistent with it ($|\mathcal{G}(\gamma^\Phi)| > 0$), and whether it is unique ($|\mathcal{G}(\gamma^\Phi)| = 1$). In the non-unique case, we can ask what the structure of $\mathcal{G}(\gamma^\Phi)$ is, and what the measures in it represent. We can also ask whether Gibbs measures with specific properties exist. The theory of Gibbs measures addresses these questions. In this paper we apply it to the case of Gibbsian specifications defined by MLNs.

3 Infinite MLNs

3.1 Definition

A Markov logic network (MLN) is a set of weighted first-order formulas. As we saw in the previous section, these can be converted to equivalent formulas in prenex CNF. We will assume throughout that all existentially quantified variables have finite domains. While this is a significant restriction, it still includes essentially all previous probabilistic relational representations as special cases. Existentially quantified formulas can now be replaced by finite disjunctions. By distributing conjunctions over disjunctions, every prenex CNF can now be converted to a quantifier-free CNF, with all variables implicitly universally quantified.

The Herbrand universe $\mathbf{U}(\mathbf{L})$ of an MLN $\mathbf{L}$ is the set of all ground terms constructible from the constants and function symbols in the MLN. The Herbrand base $\mathbf{B}(\mathbf{L})$ of $\mathbf{L}$ is the set of all ground atoms and clauses constructible from the predicates in $\mathbf{L}$, the clauses in the CNF form of $\mathbf{L}$, and the terms in $\mathbf{U}(\mathbf{L})$, replacing typed variables only by terms of the corresponding type. We assume Herbrand interpretations throughout, unless different terms are explicitly stated to represent the same object (e.g., $\forall \mathtt{x}\, \mathtt{left}(\mathtt{up}(\mathtt{x})) = \mathtt{up}(\mathtt{left}(\mathtt{x}))$). We call these *quasi-Herbrand* interpretations. We are now ready to formally define MLNs.

Definition 3. *A* Markov logic network (MLN) *$\mathbf{L}$ is a (finite) set of pairs (F_i, w_i), where F_i is a formula in first-order logic and w_i is a real number. $\mathbf{L}$ defines a countable set of variables $\mathbf{S}$ and interaction potential $\Phi^\mathbf{L} = (\Phi^\mathbf{L}_\mathbf{X})_{\mathbf{X}\in\mathcal{X}}$, $\mathcal{X}$ being the set of all finite subsets of $\mathbf{S}$, as follows:*

1. *$\mathbf{S}$ contains a binary variable for each set of atoms in $\mathbf{B}(\mathbf{L})$ whose corresponding arguments represent the same objects. The value of this variable is 1 if these (equivalent) atoms are true, and 0 otherwise.*

2. $\Phi^{\mathbf{L}}_{\mathbf{X}}(\mathbf{x}) = \sum_j w_j f_j(\mathbf{x})$*, where the sum is over the clauses* C_j *in* $\mathbf{B}(\mathbf{L})$ *whose arguments are exactly the elements of* $\mathbf{X}$*. If* $F_{i(j)}$ *is the formula in* $\mathbf{L}$ *from which* C_j *originated, and* $F_{i(j)}$ *gave rise to* n *clauses in the CNF form of* $\mathbf{L}$*, then* $w_j = w_i/n$*.* $f_j(\mathbf{x}) = 1$ *if* C_j *is true in world* $\mathbf{x}$*, and* $f_j = 0$ *otherwise.*

Let the *neighbors* $\mathbf{N}(X)$ of a ground atom X be the atoms that appear with it in some ground clause. For $\Phi^{\mathbf{L}}$ to correspond to a well-defined Gibbsian specification, it suffices that, for all $X \in \mathbf{B}(\mathbf{L})$, $|\mathbf{N}(X)| < \infty$. This ensures that the sum in Equation 2 remains finite. A sufficient condition for this is that all clauses be σ*-determinate*.

Definition 4. *A clause is* σ-determinate *if all the variables with infinite domains it contains appear in all literals.*[3] *An MLN is* σ*-determinate if all its clauses are* σ*-determinate.*

Notice that this definition does not require that all literals have the same infinite arguments; for example, the clause $\mathtt{Q}(\mathtt{x},\mathtt{y}) \Rightarrow \mathtt{R}(\mathtt{f}(\mathtt{x}),\mathtt{g}(\mathtt{x},\mathtt{y}))$ is σ-determinate. In essence, σ-determinacy requires that the neighbors of an atom be defined by functions of its arguments. Because functions can be composed indefinitely, the network can be infinite; because first-order clauses have finite length, σ-determinacy ensures that neighborhoods are still finite.

A σ-determinate MLN defines a Gibbsian specification. Given such an MLN $\mathbf{L}$, the distribution $\gamma^{\mathbf{L}}_{\mathbf{X}}$ of a set of variables $\mathbf{X} \in \mathcal{X}$ conditioned on its complement $\mathbf{S_X}$ is given by

$$\gamma^{\mathbf{L}}_{\mathbf{X}}(\mathbf{X}{=}\mathbf{x}|\mathbf{S_X}{=}\mathbf{y}) = \frac{\exp\left(\sum_j w_j f_j(\mathbf{x},\mathbf{y})\right)}{\sum_{\mathbf{x}'\in \mathrm{Dom}(\mathbf{X})}\exp\left(\sum_j w_j f_j(\mathbf{x}',\mathbf{y})\right)} \quad (5)$$

where the sum is over the clauses in $\mathbf{B}(\mathbf{L})$ that contain at least one element of $\mathbf{X}$, and $f_j(\mathbf{x},\mathbf{y}) = 1$ if clause C_j is true under the assignment $(\mathbf{x},\mathbf{y})$ and 0 otherwise. The corresponding Gibbsian specification is denoted by $\gamma^{\mathbf{L}}$.

If the MLN contains no function symbols, Definition 3 reduces to the one in Richardson and Domingos (2006), with C being the constants appearing in the MLN. This can be easily seen by substituting $\mathbf{X} = \mathbf{S}$ in Equation 5. Notice it would be equally possible to define features for conjunctions of clauses, and this may be preferable for some applications.

3.2 Existence

Let $\mathbf{L}$ be a σ-determinate MLN. The focus of this section is to show that its specification $\gamma^{\mathbf{L}}$ always admits some measure μ. It is useful to first gain some intuition as to why this might not always be the case. Consider a game where a person chooses a random natural number. Our goal is to assign probabilities to the corresponding events. This game can be expressed as a very simple MLN, containing the two formulas $\exists \mathtt{n}\ \mathtt{Chosen}(\mathtt{n})$ and $\forall \mathtt{n},\mathtt{n}'\ \mathtt{Chosen}(\mathtt{n}) \wedge \mathtt{Chosen}(\mathtt{n}') \Rightarrow \mathtt{n} = \mathtt{n}'$. Let μ be a measure admitted by the corresponding specification, and let ω_n denote the event that n is the chosen number. Then, in the limit of infinite weights, one would expect that $\mu(\bigcup \omega_n) = 1$, but in fact $\mu(\bigcup \omega_n) = \sum \mu(\omega_n) = 0$. The first equality holds because the ω_n's are disjoint, and the second one because each ω_n has zero probability of occurring by itself. But by definition $\mu(\Omega) = \mu(\bigcup \omega_n) = 1$, so there is a contradiction, and the specification does not admit a measure.[4] The reason the MLN above fails to have a measure is that the formulas are not local, in the sense that the truth value of an atom depends on the truth values of infinite others. Locality is in fact the key property for the existence of a consistent measure, and σ-determinacy ensures it (in addition to ensuring that the specification is Gibbsian).

Formally, a function $f : \Omega \to \mathbb{R}$ is said to be *local* if f only depends on a finite subset $\mathbf{V} \in \mathcal{X}$. In fact, the weaker property of *quasilocality* suffices for the existence of a Gibbs measure (Georgii, 1988).

Definition 5. *A function* $f : \Omega \to \mathbb{R}$ *is* quasilocal *if there exists a sequence* $(f_n)_{n\geq 1}$ *of local functions* f_n *such that* $\lim_{n\to\infty} \|f - f_n\| = 0$*, where* $\|.\|$ *is sup-norm. A Gibbsian specification* $\gamma = (\gamma_{\mathbf{X}})_{\mathbf{X}\in\mathcal{X}}$ *is quasilocal if each* $\gamma_{\mathbf{X}}$ *is quasilocal.*

Lemma. *Let* $\mathbf{L}$ *be a* σ*-determinate MLN, and* $\gamma^{\mathbf{L}}$ *the corresponding specification. Then* $\gamma^{\mathbf{L}}$ *is quasilocal.*

Proof. Each Hamiltonian $H^{\mathbf{L}}_{\mathbf{X}}$ is local (and hence quasilocal), since by σ-determinacy it depends only on a finite number of potentials $\phi^{\mathbf{L}}_{\mathbf{V}}$. By Proposition 2.24(b) in Georgii (1988), since all the Hamiltonians are quasilocal, the corresponding specification $\gamma^{\mathbf{L}}$ is also quasilocal. □

We now state the theorem for the existence of a measure admitted by $\gamma^{\mathbf{L}}$.

Theorem 1. *Let* $\mathbf{L}$ *be a* σ*-determinate MLN, and* $\gamma^{\mathbf{L}} = (\gamma^{\mathbf{L}}_{\mathbf{X}})_{\mathbf{X}\in\mathcal{X}}$ *be the corresponding Gibbsian specification. Then there exists a measure* μ *over* $(\Omega, \mathcal{E})$ *admitted by* $\gamma^{\mathbf{L}}$*, i.e.,* $|\mathcal{G}(\gamma^{\mathbf{L}})| \geq 1$.

Proof. To show the existence of a measure μ, we need to prove the following two conditions:

1. The net $(\gamma^{\mathbf{L}}_{\mathbf{X}}(\mathbf{X}|\mathbf{S_X}))_{\mathbf{X}\in\mathcal{X}}$ has a cluster point with respect to the weak topology on $(\Omega, \mathcal{E})$.

[3] This is related to the notion of a *determinate clause* in logic programming. In a determinate clause, the grounding of the variables in the head determines the grounding of all the variables in the body. In infinite MLNs, any literal in a clause can be inferred from the others, not just the head from the body, so we require that the (infinite-domain) variables in each literal determine the variables in the others.

[4] See Example 4.16 in Georgii (1988) for a detailed proof.

2. Each cluster point of $(\gamma^{\mathbf{L}}_{\mathbf{X}}(\mathbf{X}|\mathbf{S}_{\mathbf{X}}))_{\mathbf{X}\in\mathcal{X}}$ belongs to $\mathcal{G}(\gamma^{\mathbf{L}})$.

It is a well known result that, if all the variables X_i have finite domains, then the net in Condition 1 has a cluster point (see Section 4.2 in Georgii (1988)). Thus, since all the variables in the MLN are binary, Condition 1 holds. Further, since $\gamma^{\mathbf{L}}$ is quasilocal, every cluster point μ of the net $(\gamma^{\mathbf{L}}_{\mathbf{X}}(\mathbf{X}|\mathbf{S}_{\mathbf{X}}))_{\mathbf{X}\in\mathcal{X}}$ belongs to $\mathcal{G}(\gamma^{\mathbf{L}})$ (Comment 4.18 in Georgii (1988)). Therefore, Condition 2 is also satisfied. Hence there exists a measure μ consistent with the specification $\gamma^{\mathbf{L}}$, as required. □

3.3 Uniqueness

This section addresses the question of under what conditions an MLN admits a unique measure. Let us first gain some intuition as to why an MLN might admit more than one measure. The only condition an MLN $\mathbf{L}$ imposes on a measure is that it should be consistent with the local conditional distributions $\gamma^{\mathbf{L}}_{\mathbf{X}}$. But since these distributions are local, they do not determine the behavior of the measure at infinity. Consider, for example, an infinite two-dimensional lattice, where neighboring sites are more likely to have the same truth value than not. This can be represented by formulas of the form $\forall \mathtt{x}\ \mathtt{S}(\mathtt{x}) \Rightarrow \mathtt{S}(n(\mathtt{x}))$, where $n \in \{\mathtt{up}, \mathtt{down}, \mathtt{left}, \mathtt{right}\}$. The higher the weight w of these formulas, the more likely neighbors are to have the same value. This MLN has two extreme states: one where $\forall \mathtt{x}\ \mathtt{S}(\mathtt{x})$, and one where $\forall \mathtt{x}\ \neg\mathtt{S}(\mathtt{x})$. Let us call these states ω and $\omega_\neg$, and let ω' be a local perturbation of ω (i.e., ω' differs from ω on only a finite number of sites). If we draw a contour around the sites where ω' and ω differ, then the log odds of ω and ω' increase with wd, where d is the length of the contour. Thus long contours are improbable, and there is a measure $\mu \to \delta_\omega$ as $w \to \infty$. Since, by the same reasoning, there is a measure $\mu_\neg \to \delta_{\omega_\neg}$ as $w \to \infty$, the MLN admits more than one measure.[5]

Let us now turn to the mathematical conditions for the existence of a unique measure for a given MLN $\mathbf{L}$. Clearly, in the limit of all non-unit clause weights going to zero, $\mathbf{L}$ defines a unique distribution. Thus, by a continuity argument, one would expect the same to be true for small enough weights. This is indeed the case. To make it precise, let us first define the notion of the oscillation of a function. Given a function $f : \mathbf{X} \to \mathbb{R}$, let the oscillation of f, $\delta(f)$, be defined as

$$\begin{aligned}\delta(f) &= \max_{\mathbf{x},\mathbf{x}'\in \mathrm{Dom}(\mathbf{X})} |f(\mathbf{x}) - f(\mathbf{x}')| \\ &= \max_{\mathbf{x}} |f(\mathbf{x})| - \min_{\mathbf{x}} |f(\mathbf{x})| \qquad (6)\end{aligned}$$

[5]Notice that this argument fails for a one-dimensional lattice (equivalent to a Markov chain), since in this case an arbitrarily large number of sites can be separated from the rest by a contour of length 2. Non-uniqueness (corresponding to a non-ergodic chain) can then only be obtained by making some weights infinite (corresponding to zero transition probabilities).

The oscillation of a function is thus simply the difference between its extreme values. We can now state a sufficient condition for the existence of a unique measure.

Theorem 2. *Let $\mathbf{L}$ be a σ-determinate MLN with interaction potential Φ^L and Gibbsian specification $\gamma^{\mathbf{L}}$ such that*

$$\sup_{X_i\in\mathbf{S}} \sum_{C_j\in\mathbf{C}(X_i)} (|C_j| - 1)|w_j| < 2 \qquad (7)$$

where $\mathbf{C}(X_i)$ is the set of ground clauses in which X_i appears, $|C_j|$ is the number of ground atoms appearing in clause C_j, and w_j is its weight. Then $\gamma^{\mathbf{L}}$ admits a unique Gibbs measure.

Proof. Based on Theorem 8.7 and Proposition 8.8 in Georgii (1988), a sufficient condition for uniqueness is

$$\sup_{X_i\in\mathbf{S}} \sum_{\mathbf{V}\ni X_i} (|\mathbf{V}| - 1)\delta(\Phi^{\mathbf{L}}_{\mathbf{V}}) < 2 \qquad (8)$$

Rewriting this condition in terms of the ground formulas in which a variable X_i appears (see Definition 3) yields the desired result. □

Note that, as alluded to before, the above condition does not depend on the weight of the unit clauses. This is because for a unit clause $|C_j| - 1 = 0$. If we define the interaction between two variables as the sum of the weights of all the ground clauses in which they appear together, then the above theorem states that the total sum of the interactions of any variable with its neighbors should be less than 2 for the measure to be unique.

Two other sufficient conditions are worth mentioning briefly. One is that, if the weights of the unit clauses are sufficiently large compared to the weights of the non-unit ones, the measure is unique. Intuitively, the unit terms “drown out” the interactions, rendering the variables approximately independent. The other condition is that, if the MLN is a one-dimensional lattice, it suffices that the total interaction between the variables to the left and right of any arc be finite. This corresponds to the ergodicity condition for a Markov chain.

3.4 Non-unique MLNs

At first sight, it might appear that non-uniqueness is an undesirable property, and non-unique MLNs are not an interesting object of study. However, the non-unique case is in fact quite important, because many phenomena of interest are represented by MLNs with non-unique measures (for example, very large social networks with strong word-of-mouth effects). The question of what these measures represent, and how they relate to each other, then becomes important. This is the subject of this section.

The first observation is that the set of all Gibbs measures $\mathcal{G}(\gamma^{\mathbf{L}})$ is convex. That is, if $\mu, \mu' \in \mathcal{G}(\gamma^{\mathbf{L}})$ then $\nu \in \mathcal{G}(\gamma^{L})$, where $\nu = s\mu + (1-s)\mu', s \in (0,1)$. This is easily verified by substituting ν in Equation 4. Hence, the non-uniqueness of a Gibbs measure implies the existence of infinitely many consistent Gibbs measures. Further, many properties of the set $\mathcal{G}(\gamma^{\mathbf{L}})$ depend on the set of extreme Gibbs measures $\mathrm{ex}\,\mathcal{G}(\gamma^{\mathbf{L}})$, where $\mu \in \mathrm{ex}\,\mathcal{G}(\gamma^{\mathbf{L}})$ if $\mu \in \mathcal{G}(\gamma^{\mathbf{L}})$ cannot be written as a linear combination of two distinct measures in $\mathcal{G}(\gamma^{\mathbf{L}})$.

An important notion to understand the properties of extreme Gibbs measures is the notion of a tail event. Consider a subset $\mathbf{S}'$ of $\mathbf{S}$. Let $\sigma(\mathbf{S}')$ denote the σ-algebra generated by the set of basic events involving only variables in $\mathbf{S}'$. Then we define the tail σ-algebra $\mathcal{T}$ as

$$\mathcal{T} = \bigcap_{\mathbf{X} \in \mathcal{X}} \sigma(\mathbf{S_X}) \tag{9}$$

Any event belonging to $\mathcal{T}$ is called a tail event. $\mathcal{T}$ is precisely the set of events which do not depend on the value of any finite set of variables, but rather only on the behavior at infinity. For example, in the infinite tosses of a coin, the event that ten consecutive heads come out infinitely many times is a tail event. Similarly, in the lattice example in the previous section, the event that a finite number of variables have the value 1 is a tail event. Events in $\mathcal{T}$ can be thought of as representing macroscopic properties of the system being modeled.

Definition 6. *A measure μ is trivial on a σ-algebra $\mathcal{E}$ if $\mu(E) = 0$ or 1 for all $E \in \mathcal{E}$.*

The following theorem (adapted from Theorem 7.8 in Georgii (1988)) describes the relationship between the extreme Gibbs measures and the tail σ-algebra.

Theorem 3. *Let $\mathbf{L}$ be a σ-determinate MLN, and $\gamma^{\mathbf{L}}$ denote the corresponding Gibbsian specification. Then the following results hold:*

1. *A measure $\mu \in \mathrm{ex}\,\mathcal{G}(\gamma^{\mathbf{L}}))$ iff it is trivial on the tail σ-algebra $\mathcal{T}$.*
2. *Each measure μ is uniquely determined by its behavior on the tail σ-algebra, i.e., if $\mu_1 = \mu_2$ on $\mathcal{T}$ then $\mu_1 = \mu_2$.*

It is easy to see that each extreme measure corresponds to some particular value for all the macroscopic properties of the network. In physical systems, extreme measures correspond to phases of the system (e.g., liquid vs. gas, or different directions of magnetization), and non-extreme measures correspond to probability distributions over phases. Uncertainty over phases arises when our knowledge of a system is not sufficient to determine its macroscopic state. Clearly, the study of non-unique MLNs beyond the highly regular ones statistical physicists have focused on promises to be quite interesting. In the next section we take a step in this direction by considering the problem of satisfiability in the context of MLN measures.

3.5 Infinite Weight Limit

This section examines the properties of measures over an MLN in the limit of all equal infinite weights. As we will see, this limiting case is central to relationship between measures over MLNs and the problem of satisfiability.

Consider an MLN $\mathbf{L}$ such that each clause in its CNF form has the same weight w. In the limit $w \to \infty$, $\mathbf{L}$ does not correspond to a valid Gibbsian specification, since the Hamiltonians defined in Equation 2 are no longer finite. We will show that in the limit of all equal infinite clause weights, conditional distributions in Equation 5 are still well defined and are equi-distributions over those assignments of variables in $\mathbf{X}$ which satisfy the maximum number of clauses given $\mathbf{S_X} = \mathbf{y}$. Further, we will show that we can still talk about the existence of a measure consistent with these conditional distributions, because these distributions constitute a valid specification (though not Gibbsian) under the same conditions as in the finite weight case.

Let KB be a first-order knowledge base. Let us consider Equation 5 when we substitute $w_j = w, \forall j$. It can then be rewritten as

$$\gamma^{\mathbf{L}}_{\mathbf{X}}(\mathbf{X} = \mathbf{x} | \mathbf{S_X} = \mathbf{y}) = \frac{1}{1 + D(\mathbf{x}, \mathbf{y})} \tag{10}$$

where

$$D(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{x}' \in Dom(\mathbf{X}) / \{\mathbf{x}\}} \exp\left(w \sum_j (f_j(\mathbf{x}', \mathbf{y}) - f_j(\mathbf{x}, \mathbf{y})) \right) \tag{11}$$

This in turn can be rewritten as

$$D(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{x}' \in Dom(\mathbf{X}) / \{\mathbf{x}\}} \exp\left(w(n_{\mathbf{X}}(\mathbf{x}', \mathbf{y}) - n_{\mathbf{X}}(\mathbf{x}, \mathbf{y})) \right) \tag{12}$$

where $n_{\mathbf{X}}(\mathbf{x}, \mathbf{y})$ is the total number of ground clauses which evaluate to true given the assignment $(\mathbf{x}, \mathbf{y})$. Let $n^{max}_{\mathbf{X}}(\mathbf{y}) = max_{\mathbf{x} \in Dom(\mathbf{X})} n_{\mathbf{X}}(\mathbf{x}, \mathbf{y})$ be the maximum number of clauses which can be satisfied by any assignment $\mathbf{X} = \mathbf{x}$, given $\mathbf{S_X} = \mathbf{y}$. Then, in the limit $w \to \infty$, $D(\mathbf{x}, \mathbf{y}) = k$, k being some constant, for the $\mathbf{X}$ assignments satisfying $n^{max}_{\mathbf{X}}(\mathbf{y})$ number of clauses. We will call these the *maximal assignments* for a given value of $\mathbf{y}$. We will use the notation $\eta_{\mathbf{X}}(\mathbf{y})$ to denote the set of all the maximal assignments to $\mathbf{X}$ given $\mathbf{S_X} = \mathbf{y}$. Note that for all the non-maximal assignments, $D(\mathbf{x}, \mathbf{y}) = \infty$. For a given $\mathbf{y}$, equation 10 assigns equal probability to all the maximal assignments. In particular, if the KB is satisfiable and $\mathbf{S_X} = \mathbf{y}$ is some partial satisfying assignment, then the maximal assignments are precisely the completions of the partial satisfying assignment dictated by $\mathbf{y}$ and equation 10 defines a uniform distribution over them.

Now, we need to show that the family $\gamma^{\mathbf{L}} = (\gamma^{\mathbf{L}}_{\mathbf{X}})_{\mathbf{X} \in \mathcal{X}}$ is a valid specification. Going to the definition of a specification (Def 1.23 (Georgii, 1988)), γ is a specification [6], if for any $\mathbf{X_1}, \mathbf{X_2} \in \mathcal{X}$, $\mathbf{X_1} \subset \mathbf{X_2}$, we have

$$\gamma_{\mathbf{X_2}}(\mathbf{Z}|\mathbf{S_{X_2}}) = \sum_{\mathbf{X_{21}}=\mathbf{X_2}-\mathbf{X_1}} \gamma_{\mathbf{X_2}}(\mathbf{X_{21}}|\mathbf{S_{X_2}})\gamma_{\mathbf{X_1}}(\mathbf{Z}|\mathbf{X_{21}}, \mathbf{S_{X_2}}) \tag{13}$$

for all $\mathbf{Z} \subseteq \mathbf{X_1}$. A sufficient condition for this to be true is

$$\gamma_{\mathbf{X_2}}(\mathbf{X_1}|\mathbf{X_{21}}, \mathbf{S_{X_2}}) = \gamma_{\mathbf{X_1}}(\mathbf{X_1}|\mathbf{X_{21}}, \mathbf{S_{X_2}}) \tag{14}$$

Rewriting RHS using Bayes rule we get

$$\frac{\gamma_{\mathbf{X_2}}(\mathbf{X_1}, \mathbf{X_{21}}|\mathbf{S_{X_2}})}{\gamma_{\mathbf{X_2}}(\mathbf{X_{21}}|\mathbf{S_{X_2}})} = \gamma_{\mathbf{X_1}}(\mathbf{X_1}|\mathbf{X_{21}}, \mathbf{S_{X_2}}) \tag{15}$$

For the assignments $\mathbf{X_{21}} = \mathbf{x_{21}}$ such that denominator of RHS becomes zero, above expression is not defined. But it can easily be seen that the corresponding terms cancel out from equation 13 under such assignments to variables in $\mathbf{X_{21}}$. Hence, we can altogether ignore these assignments. Substituting $\mathbf{X_1} = \mathbf{x_1}$, $\mathbf{X_{21}} = \mathbf{x_{21}}$ and $\mathbf{S_{X_2}} = \mathbf{y_2}$ and expanding out the denominator, we get

$$\frac{\gamma_{\mathbf{X_2}}(\mathbf{x_1}, \mathbf{x_{21}}|\mathbf{y_2})}{\sum_{\mathbf{x'_1} \in \mathbf{Dom}(\mathbf{X_1})} \gamma_{\mathbf{X_2}}(\mathbf{x'_1}, \mathbf{x_{21}}|\mathbf{y_2})} = \gamma_{\mathbf{X_1}}(\mathbf{x_1}|\mathbf{x_{21}}, \mathbf{y_2}) \tag{16}$$

Let us first consider the case when the assignment $(\mathbf{x_1}, \mathbf{x_{21}})$ is maximal given $\mathbf{y_2}$. Let $\mathbf{y_1} = (\mathbf{x_{21}}, \mathbf{y_2})$ denote the assignment to the variables in $\mathbf{S_{X_1}}$. Then, above equation can be written as

$$\frac{1/|\eta_{\mathbf{X_2}}(\mathbf{y_2})|}{\sum_{\mathbf{x_1} \in \eta_{\mathbf{X_1}}(\mathbf{y_1})} 1/|\eta_{\mathbf{X_2}}(\mathbf{y_2})|} = 1/|\eta_{\mathbf{X_1}}(\mathbf{y_1})| \tag{17}$$

or equivalently as

$$\frac{1/|\eta_{\mathbf{X_2}}(\mathbf{y_2})|}{|\eta_{\mathbf{X_1}}(\mathbf{y_1})|/|\eta_{\mathbf{X_2}}(\mathbf{y_2})|} = 1/|\eta_{\mathbf{X_1}}(\mathbf{y_1})| \tag{18}$$

which is clearly true. The case when $(\mathbf{x_1}, \mathbf{x_{21}})$ is not maximal given $\mathbf{y_2}$ can be split into two sub-cases. The first is when when $\mathbf{x_{21}}$ can not be in any maximal assignment given $\mathbf{y_2}$. In this case, the denominator of RHS in equation 16 becomes zero. But it has already been considered. The second is when $\mathbf{x_1}$ is not maximal given $(\mathbf{x_{21}}, \mathbf{y_2})$. Then, both RHS and LHS are zero in equation 16. Hence, $\gamma^{\mathbf{L}}$ defines a valid specification in the limit of all equal infinite weights.

Let $\mathbf{L}_\infty$ be a σ-determinate MLN in the limit of all equal infinite weights. Let $\gamma^{\mathbf{L}_\infty}$ denote the corresponding specification as defined above. It is easy to see that we can still apply Theorem 1, as is, to show the existence of a measure consistent with $\gamma^{\mathbf{L}_\infty}$.

[6]We will drop $\mathbf{L}$ in the superscript for clarity

4 Satisfiability

Richardson and Domingos (2006) showed that, in finite domains, first-order logic can be viewed as the limiting case of Markov logic when all weights tend to infinity, in the following sense. If we convert a satisfiable knowledge base $\mathbf{K}$ into an MLN $\mathbf{L_K}$ by assigning the same weight $w \to \infty$ to all clauses, then $\mathbf{L_K}$ defines a uniform distribution over the worlds satisfying $\mathbf{K}$. Further, $\mathbf{K}$ entails a formula α iff $\mathbf{L_K}$ assigns probability 1 to the set of worlds satisfying α (Proposition 4.3). We would like to extend this result to infinite domains. We will first define the notion of a *satisfying measure*, which is central to the results presented in this section.

Definition 7. *Let* $\mathbf{L}$ *be a* σ*-determinate MLN. Given a clause* $C_i \in \mathbf{B}(\mathbf{L})$*, let* $\mathbf{V}_i$ *denote the set of variables appearing in* C_i*. A measure* $\mu \in \mathcal{G}(\gamma^{\mathbf{L}})$ *is said to be a* satisfying measure *for* $\mathbf{L}$ *if, for every ground clause* $C_i \in \mathbf{B}(\mathbf{L})$*,* μ *assigns non-zero probability only to the satisfying assignments of the variables in* C_i*, i.e.,* $\mu(\mathbf{V}_i = \mathbf{v}_i) > 0$ *implies that* $\mathbf{V}_i = \mathbf{v}_i$ *is a satisfying assignment for* C_i*.* $\mathcal{S}(\gamma^{\mathbf{L}})$ *denotes the set of all satisfying measures for* $\mathbf{L}$*.*

Informally, a satisfying measure assigns non-zero probability only to those worlds which are consistent with all the formulas in $\mathbf{L}$. Intuitively, existence of a satisfying measure for $\mathbf{L}$ should be in some way related to the existence of a satisfying assignment for the corresponding knowledge base. Our next theorem formalizes this intuition.

Theorem 4. *Let* $\mathbf{K}$ *be a knowledge base all of whose clauses are* σ*-determinate, and let* $\mathbf{L}_\infty$ *be the MLN obtained by assigning weight* $w \to \infty$ *to all the clauses in* $\mathbf{K}$*. Then there exists a satisfying measure for* $\mathbf{L}_\infty$ *iff* $\mathbf{K}$ *is satisfiable. Mathematically,*

$$|\mathcal{S}(\gamma^{\mathbf{L}_\infty})| > 0 \Leftrightarrow \mathit{Satisfiable}(\mathbf{K}) \tag{19}$$

Proof. Let us first prove that existence of a satisfying measure implies satisfiability of $\mathbf{K}$. This is equivalent to proving that unsatisfiability of $\mathbf{K}$ implies non-existence of a satisfying measure. Let $\mathbf{K}$ be unsatisfiable. Equivalently, $\mathbf{B}(\mathbf{K})$, the Herbrand base of $\mathbf{K}$, is unsatisfiable. By Herbrand's theorem, there exists a finite set of ground clauses $\mathbf{C} \subseteq \mathbf{B}(\mathbf{K})$ that is unsatisfiable. Let $\mathbf{V}$ denote the set of variables appearing in $\mathbf{C}$. Then every assignment $\mathbf{v}$ to the variables in $\mathbf{V}$ violates some clause in $\mathbf{C}$. Let μ denote a measure for $\mathbf{L}_\infty$. Since μ is a probability measure, $\sum_{\mathbf{v} \in \mathrm{Dom}(\mathbf{V})} \mu(\mathbf{V} = \mathbf{v}) = 1$. Further, since $\mathbf{V}$ is finite, there exists some $\mathbf{v} \in \mathrm{Dom}(\mathbf{V})$ such that $\mu(\mathbf{V} = \mathbf{v}) > 0$. Let $C_i \in \mathbf{C}$ be some clause violated by the assignment $\mathbf{v}$ (every assignment violates some clause). Let $\mathbf{V}_i$ denote the set of variables in C_i and $\mathbf{v}_i$ be the restriction of assignment $\mathbf{v}$ to the variables in $\mathbf{V}_i$. Then $\mathbf{v}_i$ is an unsatisfying assignment for C_i. Further, $\mu(\mathbf{V}_i = \mathbf{v}_i) \geq \mu(\mathbf{V} = \mathbf{v}) > 0$. Hence μ cannot be a satisfying measure for $\mathbf{L}_\infty$. Since the above argument holds for any $\mu \in \mathcal{G}(\gamma^{\mathbf{L}_\infty})$, there does not exist a satisfying measure for $\mathbf{L}_\infty$ when $\mathbf{K}$ is unsatisfiable.

Next, we need to prove that satisfiability of $\mathbf{K}$ implies existence of a satisfying measure. For the clarity of notation, we will drop the superscript from $\gamma^{\mathbf{L}_\infty}$ in the following part of the proof. Our proof builds on the proof for the existence of a measure presented in Pfeffer (2000) (see Chapter 7).

Consider a finite set of variables $\mathbf{X}$. An assignment $\mathbf{X} = \mathbf{x}$ is called a *partial satisfying assignment* to the variables in $\mathbf{X}$ if $\exists \omega \in \Omega$, such that $\omega_{\mathbf{X}} = \mathbf{x}$ and ω is a satisfying assignment. Let $p_{\mathbf{X}}$ denote a probability distribution over $\mathbf{X}$. We call $p_{\mathbf{X}}$ a *satisfying probability distribution* if it assigns non-zero probability only to to partial satisfying assignments of $\mathbf{X}$. Let $\Delta_{\mathbf{X}}$ denote the set all satisfying probability distributions over $\mathbf{X}$. It is a subset of R^n, where n is the number of values in $Dom(\mathbf{X})$. $\Delta_{\mathbf{X}}$ consists of points $p \in R^n$ satisfying the constraints that $p_i \geq 0$, $p_i = 0$ if i^{th} assignment is not a partial satisfying assignment and $\sum_{i=1}^{n} p_i = 1$. $\Delta_{\mathbf{X}}$ is a closed and bounded, and therefore compact, subset of R^n. Further, since K is satisfiable, $\Delta_{\mathbf{X}}$ is non-empty for all $\mathbf{X}$.

Let $\mathbf{X}$ be a finite set of variables, $\mathbf{Y}$ the neighbors of $\mathbf{X}$. It is easy to see that $\forall \mathbf{X}' \supseteq \mathbf{X}$, we have

$$\gamma_{\mathbf{X}'}(\mathbf{X}|\mathbf{S_X}) = \gamma_{\mathbf{X}'}(\mathbf{X}|\mathbf{Y}) \tag{20}$$

In other words, any finite set of variables is independent of the rest of the network given its neighbors. We define the mapping $F_{\mathbf{X}}$ from $\Delta_{\mathbf{Y}}$ to $\Delta_{\mathbf{X}}$ as follows. For each $q \in \Delta_{\mathbf{Y}}$,

$$F_{\mathbf{X}}(q) = \sum_{\mathbf{y} \in Dom(\mathbf{Y})} q(\mathbf{y}) \gamma_{\mathbf{X}}(\mathbf{X}|\mathbf{Y} = \mathbf{y}) \tag{21}$$

For above mapping to be valid, we need to show that if $q \in \Delta_{\mathbf{Y}}$ then $p = F_{\mathbf{X}}(q) \in \Delta_{\mathbf{X}}$. Clearly, p is a distribution over $\mathbf{X}$. Hence, we only need to show that p is a satisfying distribution. Consider a non-satisfying partial assignment $\mathbf{x}$ to the variables in $\mathbf{X}$. Then, each term in the summation above is zero either because a) $\mathbf{y}$ is not a partial satisfying assignment (and hence $q(\mathbf{y}) = 0$) or b) $\mathbf{y}$ is a partial satisfying assignment and $\mathbf{X} = \mathbf{x}$ is not maximal given $\mathbf{Y} = \mathbf{y}$ (and hence $\gamma_{\mathbf{X}}(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y}) = 0$).

Lemma. *Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be two finite set of variables, with $\mathbf{X}_1 \subseteq \mathbf{X}_2$, and let $\mathbf{Y}_1$ and $\mathbf{Y}_2$ be neighbors of $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively. If $p \in \Delta_{\mathbf{X}_2}$ is in the image of $\Delta_{\mathbf{Y}_2}$ under $F_{\mathbf{X}_2}$, then $\sum_{\mathbf{X}_2 - \mathbf{X}_1} p$ is in the image of $\Delta_{\mathbf{Y}_1}$ under $F_{\mathbf{X}_1}$.*

Proof. First note that $\mathbf{Y_1} - \mathbf{Y_2} \subseteq \mathbf{X_2}$. This means that all the neighbors of $\mathbf{Y_1} - \mathbf{Y_2}$ lie in the set $\mathbf{Y_2}$. Using Equation 20, Equation 14 can be rewritten as

$$\gamma_{\mathbf{X_2}}(\mathbf{X_1}|\mathbf{Y_1}, \mathbf{Z}) = \gamma_{\mathbf{X_1}}(\mathbf{X_1}|\mathbf{Y_1}, \mathbf{Z}) \tag{22}$$

where $\mathbf{Z}$ is some finite set of variables.

If $p \in \Delta_{\mathbf{X}_2}$ is in the image of $\Delta_{\mathbf{Y}_2}$ under $F_{\mathbf{X}_2}$, there is some $q \in \Delta_{\mathbf{Y}_2}$ such that

$$p(\mathbf{X}_2) = \sum_{\mathbf{y}_2} q(\mathbf{y_2}) \gamma_{\mathbf{X}_2}(\mathbf{X_2}|\mathbf{Y_2} = \mathbf{y_2}) \tag{23}$$

Let p' be $\sum_{\mathbf{X}_2 - \mathbf{X}_1} p$. We have

$$p'(\mathbf{x_1}) = \sum_{\mathbf{x_{21}} \in Dom(\mathbf{X_2} - \mathbf{X_1})} p(\mathbf{x_1}, \mathbf{x_{21}}) \tag{24}$$

$$\begin{aligned}
&= \sum_{\mathbf{x_{21}}} \sum_{\mathbf{y_2}} q(\mathbf{y_2}) \gamma_{\mathbf{X_2}}(\mathbf{x_1}, \mathbf{x_{21}}|\mathbf{y_2}) \\
&= \sum_{\mathbf{y_2}} q(\mathbf{y_2}) \sum_{\mathbf{x_{21}}} \gamma_{\mathbf{X_2}}(\mathbf{x_1}, \mathbf{x_{21}}|\mathbf{y_2}) \\
&= \sum_{\mathbf{y_2}} q(\mathbf{y_2}) \gamma_{\mathbf{X_2}}(\mathbf{x_1}|\mathbf{y_2}) \\
&= \sum_{\mathbf{y_2}} q(\mathbf{y_2}) \sum_{\mathbf{y_{12}} \in Dom(\mathbf{Y_1} - \mathbf{Y_2})} \gamma_{\mathbf{X_2}}(\mathbf{x_1}, \mathbf{y_{12}}|\mathbf{y_2}) \\
&= \sum_{\mathbf{y_2}} q(\mathbf{y_2}) \sum_{\mathbf{y_{12}}} \gamma_{\mathbf{X_2}}(\mathbf{x_1}|\mathbf{y_{12}}, \mathbf{y_2}) \gamma_{\mathbf{X_2}}(\mathbf{y_{12}}|\mathbf{y_2}) \\
&\quad (using\ Equation\ 22) \\
&= \sum_{\mathbf{y_2}} q(\mathbf{y_2}) \sum_{\mathbf{y_{12}}} \gamma_{\mathbf{X_1}}(\mathbf{x_1}|\mathbf{y_{12}}, \mathbf{y_2}) \gamma_{\mathbf{X_2}}(\mathbf{y_{12}}|\mathbf{y_2}) \\
&= \sum_{\mathbf{y_2}} \sum_{\mathbf{y_{12}}} q(\mathbf{y_2}) \gamma_{\mathbf{X_2}}(\mathbf{y_{12}}|\mathbf{y_2}) \gamma_{\mathbf{X_1}}(\mathbf{x_1}|\mathbf{y_{12}}, \mathbf{y_2}) \\
&= \sum_{\mathbf{y_1}} \sum_{\mathbf{y_{21}} \in \mathbf{Dom}(\mathbf{Y_2} - \mathbf{Y_1})} q(\mathbf{y_2}) \gamma_{\mathbf{X_2}}(\mathbf{y_{12}}|\mathbf{y_2}) \gamma_{\mathbf{X_1}}(\mathbf{x_1}|\mathbf{y_1}) \\
&= \sum_{\mathbf{y_1}} \left[\sum_{\mathbf{y_{21}}} q(\mathbf{y_2}) \gamma_{\mathbf{X_2}}(\mathbf{y_{12}}|\mathbf{y_2}) \right] \gamma_{\mathbf{X_1}}(\mathbf{x_1}|\mathbf{y_1})
\end{aligned} \tag{25}$$

Note that though $\mathbf{y_2}$ and $\mathbf{y_{12}}$ do not appear explicitly as summation variables in the last two lines of above expression, their values can be uniquely determined given $\mathbf{y_1}$ and $\mathbf{y_{21}}$. Let

$$q'(\mathbf{y_1}) = \sum_{\mathbf{y_{21}}} q(\mathbf{y_2}) \gamma_{\mathbf{X_2}}(\mathbf{y_{12}}|\mathbf{y_2}) \tag{26}$$

Now, if we show that $q' \in \Delta_{\mathbf{Y_1}}$, then $p' = F_{\mathbf{X_1}}(q')$ which is precisely what we set out to prove. So, only thing which remains to be shown is $q' \in \Delta_{\mathbf{Y_1}}$. Now,

$$\begin{aligned}
q'(\mathbf{y_1}) &= \sum_{\mathbf{y_{21}}} q(\mathbf{y_2}) \gamma_{\mathbf{X_2}}(\mathbf{y_{12}}|\mathbf{y_2}) \\
&= \sum_{\mathbf{y_{21}}} r(\mathbf{y_2}, \mathbf{y_{12}}) \\
&= \sum_{\mathbf{y_{21}}} r(\mathbf{y_1}, \mathbf{y_{21}})
\end{aligned} \tag{27}$$

where r is a distribution over $(\mathbf{y_1}, \mathbf{y_{21}})$. Hence, q' is a distribution over $\mathbf{Y_1}$. Next, we need to show that q' is a satisfying distribution. Let $\mathbf{y_1}$ be a non-satisfying partial assignment to the variables in $\mathbf{Y_1}$. Then, considering Equation 26, for any assignment $\mathbf{y_{21}}$, either a) $\mathbf{y_2}$ is non-satisfying assignment to variables in $\mathbf{Y_2}$ in which case $q(\mathbf{y_2}) = 0$ or b) $\mathbf{y_{12}}$ is not maximal given $\mathbf{y_2}$ in which case $\gamma_{\mathbf{X_2}}(\mathbf{y_{12}}|\mathbf{y_2}) = 0$. This implies that $q'(\mathbf{y_1}) = 0$ for any non-satisfying partial assignment $\mathbf{y_1}$. Hence, q' is a satisfying distribution. Therefore p' is in the image of $\Delta_{\mathbf{X}_1}$ under $F_{\mathbf{X}_1}$, as required.

□

Now, we will show the existence of the required satisfying measure by construction. We begin by constructing an increasing sequence of finite sets that covers all the variables of $\mathbf{S}$. Since the number of variables in the network is countably infinite, we can order the variables is some sequence $\{X_1, X_2, \cdots X_n, \cdots\}$ such that for every variable $V \in V_M, \exists k$ satisfying $V = X_k$. For $n > 0$, define

$$\mathbf{X}_n = \{X_1, X_2, \cdots X_n\} \tag{28}$$

Note that, $\mathbf{X_1} \subseteq \mathbf{X_2} \subseteq \cdots$ by definition. Now, for each $\mathbf{X_n}$ we define a set of probability distributions over $\mathbf{X_n}$ that are consistent with the conditional distribution $\gamma_{\mathbf{X_n}}$, for some distribution over the neighbors of $\mathbf{X_n}$. Let $\mathbf{Y_n}$ be the neighbors of $\mathbf{X}_n$. Define the set $\mathbf{I_n} \subseteq \Delta_{\mathbf{X_n}}$ to be the image of $\Delta_{\mathbf{Y_n}}$ under the mapping mapping $F_{\mathbf{X_n}}$. S_n is the continuous image of the compact set $\Delta_{\mathbf{Y_n}}$, so it is compact. $\mathbf{I_n}$ is obviously also non-empty, since it is the image of a non-empty set.

Next, we use the compactness to show that for each $\mathbf{X_n}$, there is some non-empty subset of $\mathbf{I_n}$ of distributions that are consistent with some distribution in $\mathbf{I_m}$ for every $m \geq n$. For $m \geq n$, define the set $\mathbf{T_n^m} \subseteq \Delta_{\mathbf{X_n}}$ to be the image of $\mathbf{I_m}$ under the marginalization operator $\sum_{\mathbf{X_m}-\mathbf{X_n}}$. We show the following:

1. $\mathbf{T_n^m}$ is closed. This is clear, since it is the continuous image of a compact set, and therefore compact, and therefore closed.

2. $\mathbf{T_n^m}$ is non-empty. This is obvious, since it is the image of non-empty set.

3. $\mathbf{T_n^m} \subseteq \mathbf{I_n}$. This follows immediately from the previous Lemma.

4. If $m_1 < m_2$, $\mathbf{T_n^{m_1}} \supseteq \mathbf{T_n^{m_2}}$. This follows from point 3 as follows. $\mathbf{T_n^{m_2}}$ is the image of $\mathbf{I_{m_2}}$ under the marginalization operator $G_2 = \sum_{\mathbf{X_{m_2}}-\mathbf{X_n}}$, which is the composition of the operators $H = \sum_{\mathbf{X_{m_2}}-\mathbf{X_{m_1}}}$ and $G_1 = \sum_{\mathbf{X_{m_1}}-\mathbf{X_n}}$. Since the image of $\mathbf{T_n^{m_2}}$ under H is a subset of $\mathbf{I_{m_1}}$ by point 3, and $\mathbf{T_n^{m_1}}$ is the image of $\mathbf{I_{m_1}}$ under G_1, it follows that $\mathbf{T_n^{m_1}} \supseteq \mathbf{T_n^{m_2}}$.

It follows, therefore, that $(\mathbf{T_m^n})_{m=n}^{\infty}$ is a decreasing sequence of closed, non-empty subsets of $\mathbf{I_n}$. Define $\mathbf{T_n}$ to be $\cap_{m=n}^{\infty}\mathbf{T_n^m}$. By compactness of $\mathbf{I_n}$, $\mathbf{T_n}$ is a non-empty subset of $\mathbf{I_n}$.

Each $\mathbf{T_n}$ is a set of distributions over $\mathbf{X_n}$ such that for every p in $\mathbf{T_n}$, p is consistent with the conditional distribution $\gamma_{\mathbf{X_n}}$ for some distribution q over $\mathbf{Y_n}$, and furthermore, for every $m \geq n$, p is equal to $\sum_{\mathbf{X_m}-\mathbf{X_n}} p'$ for some $p' \in \mathbf{T_m}$. The final stage of the construction is to use $\mathbf{T_n}$ to define a function F over the basic events, as follows.

We construct a sequence of distributions F_n over $\mathbf{X_n}$ as follows. For F_1, choose any element of $\mathbf{T_{X_1}}$. For F_n with $n > 1$, choose an element of $\mathbf{T_{X_n}}$ such that $F_{n-1} = \sum_{X_n} F_n$. Since $F_{n-1} \in \mathbf{T_{X_{n-1}}}$, such choice must exist. For $m_1 < m_2$, we have inductively that $F_{m_1} = \sum_{\mathbf{X}_{m_2}-\mathbf{X}_{m_1}} F_{m_2}$. It follows that we can define the function F over basic events unambiguously by setting $F(E) = F_n(E)$ for any n such that all variables mentioned by $\mathbf{X}$ are in $\mathbf{X_n}$.

It is easy to see that F is additive. Let $E_1, E_2, \cdots E_m$ be a set of disjoint basic events such that their union E is a basic event. Let n be such that all variables mentioned in any of the E_i or in E are in $\mathbf{X}_n$. Then, since F_n is a probability distribution,

$$F(E) = F_n(E) = F_n(\cup_{i=1}^m E_i) = \sum_{i=1}^{m} F_n(E_m) = \sum_{i=1} F(E_i) \tag{29}$$

Using the fact that no basic event is the infinitely countable disjoint union of basic events, additivity is sufficient to ensure countable additivity. Further, Ω can be expressed as the event $[X_1 = 0 \vee X_1 = 1]$. Then, $\mu(\Omega) = F_1([X_1 = 0 \vee X_1 = 1]) = 1$. Hence, μ is a probability measure over the space $(\Omega, \mathcal{E})$.

Finally, μ satisfies the required conditions of being consistent with the local probabilistic dependencies (Definition 1.23 (Georgii, 1988)). Let $\mathbf{X}$ be a finite set of variables. Let $\mathbf{Y}$ denote the set of neighbors of $\mathbf{X}$. Let $\mathbf{Z}$ denote the set of neighbors of $\mathbf{X} \cup \mathbf{Y}$. By Bayes rule, $\mu(\mathbf{X}|\mathbf{Y}) = \frac{\mu(\mathbf{X}\cup\mathbf{Y})}{\mu(\mathbf{Y})}$. By construction, for some satisfying distribution q over $\mathbf{Z}$, we have

$$\begin{aligned}
\frac{\mu(\mathbf{X} \cup \mathbf{Y})}{\mu(\mathbf{Y})} &= \frac{\sum_{\mathbf{z}} q(\mathbf{z})\gamma_{\mathbf{X}\cup\mathbf{Y}}(\mathbf{X} \cup \mathbf{Y}|\mathbf{Z})}{\sum_{\mathbf{x}}\sum_{\mathbf{z}} q(\mathbf{z})\gamma_{\mathbf{X}\cup\mathbf{Y}}(\mathbf{X} \cup \mathbf{Y}|\mathbf{Z})} \\
&= \frac{\sum_{\mathbf{z}} q(\mathbf{z})\gamma_{\mathbf{X}\cup\mathbf{Y}}(\mathbf{X}|\mathbf{Y} \cup \mathbf{Z})\gamma_{\mathbf{X}\cup\mathbf{Y}}(\mathbf{Y}|\mathbf{Z})}{\sum_{\mathbf{x}}\sum_{\mathbf{z}} q(\mathbf{z})\gamma_{\mathbf{X}\cup\mathbf{Y}}(\mathbf{X}|\mathbf{Y} \cup \mathbf{Z})\gamma_{\mathbf{X}\cup\mathbf{Y}}(\mathbf{Y}|\mathbf{Z})} \\
&= \frac{\sum_{\mathbf{z}} q(\mathbf{z})\gamma_{\mathbf{X}}(\mathbf{X}|\mathbf{Y})\gamma_{\mathbf{X}\cup\mathbf{Y}}(\mathbf{Y}|\mathbf{Z})}{\sum_{\mathbf{x}}\sum_{\mathbf{z}} q(\mathbf{z})\gamma_{\mathbf{X}}(\mathbf{X}|\mathbf{Y})\gamma_{\mathbf{X}\cup\mathbf{Y}}(\mathbf{Y}|\mathbf{Z})} \\
&= \frac{\gamma_{\mathbf{X}}(\mathbf{X}|\mathbf{Y})\sum_{\mathbf{z}} q(\mathbf{z})\gamma_{\mathbf{X}\cup\mathbf{Y}}(\mathbf{Y}|\mathbf{Z})}{\sum_{\mathbf{x}}\gamma_{\mathbf{X}}(\mathbf{X}|\mathbf{Y})\sum_{\mathbf{z}} q(\mathbf{z})\gamma_{\mathbf{X}\cup\mathbf{Y}}(\mathbf{Y}|\mathbf{Z})} \\
&= \gamma_{\mathbf{X}}(\mathbf{X}|\mathbf{Y})
\end{aligned}$$

Therefore, μ is a measure admitted by the specification $\gamma^{\mathbf{L}_\infty}$. Further, given any finite set of variables $\mathbf{X}$, by construction μ assigns non-zero probability only to satisfying partial assignments of $\mathbf{X}$. This is in turn implies that for every clause C_i, μ assigns non-zero probability only to the satisfying assignments of variables in C_i. Hence, μ is a satisfying measure, as required. $\square$

Corollary. *Let $\mathbf{K}$ be satisfiable. Let α be a first-order formula, and $\mathbf{L}_\infty^\alpha$ be the MLN obtained by assigning weight $w \to \infty$ to all clauses in $\mathbf{K} \cup \{\neg\alpha\}$. Then $\mathbf{K}$ entails α iff $\mathbf{L}_\infty^\alpha$ has no satisfying measure. Mathematically,*

$$\mathbf{K} \models \alpha \;\Leftrightarrow\; |\mathcal{S}(\gamma^{\mathbf{L}_\infty^\alpha})| = 0 \tag{30}$$

Thus, for knowledge bases containing only σ-determinate clauses and no infinite existentials, with Herbrand or quasi-Herbrand interpretations, first-order logic can be viewed as the limiting case of Markov logic when all weights tend to infinity. Whether these conditions can be relaxed is a question for future work.

5 Related Work

A number of relational representations capable of handling infinite domains have been proposed in recent years. Generally, they rely on strong restrictions to make this possible. To our knowledge, Markov logic is the most flexible language for modeling infinite relational domains to date. In this section we briefly review the main approaches.

Stochastic logic programs (Muggleton, 1996) are generalizations of probabilistic context-free grammars, which allow for infinite derivations but as a result do not always represent valid distributions (Booth & Thompson, 1973). SLPs (and related languages like independent choice logic (Poole, 1997) and PRISM (Sato & Kameya, 1997)) are a special case of Markov logic, because logic programs are a special case of first-order logic and PCFGs are a special case of Gibbs distributions (Chi, 1999).

Many approaches combine logic programming and Bayesian networks. The most advanced one is arguably Bayesian logic programs (Kersting & De Raedt, 2001). Kersting and De Raedt show that, if all nodes have a finite number of ancestors, a BLP represents a unique distribution. This is a stronger restriction than finite neighborhoods. BLPs can be converted into Markov logic in the same way that Bayesian networks can be converted into Markov networks.

Jaeger (1998) shows that probabilistic queries are decidable for a very restricted language where a ground atom cannot depend on other groundings of the same predicate. Jaeger shows that if this restriction is removed queries become undecidable.

Recursive probability models are a combination of Bayesian networks and description logics (Pfeffer & Koller, 2000). Like Markov logic, RPMs require finite neighborhoods, and in fact existence for RPMs can be proved succinctly by converting them to Markov logic and applying Theorem 1. Pfeffer and Koller show that RPMs do not always represent unique distributions, but do not study conditions for uniqueness. Description logics are a restricted subset of first-order logic, and thus MLNs are considerably more flexible than RPMs.

Contingent Bayesian networks (Milch et al., 2005) allow infinite ancestors, but require that, for each variable with infinite ancestors, there exist a set of mutually exclusive and exhaustive contexts (assignments to finite sets of variables) such that in every context only a finite number of ancestors affect the probability of the variable. Effectively, in any given world the network is always finite. This is a strong restriction, excluding even simple infinite models like Markov chains.

Multi-entity Bayesian networks are another relational extension of Bayesian networks (Laskey & Costa, 2005). Laskey and Costa claim that MEBNs allow infinite parents and arbitrary first-order formulas, but the definition of MEBN explicitly requires that, for each atom X and increasing sequence of substates $S_1 \subset S_2 \subset \ldots$, there exist a finite N such that $P(X|S_k) = P(X|S_N)$ for $k > N$. This assumption is tantamount to requiring that nodes have a finite number of ancestors, and makes it straighforward to prove existence and uniqueness using Kolmogorov's theorem. But it necessarily excludes many dependencies expressible in first-order logic (e.g., $\forall \mathtt{x}\ \exists \mathtt{y}\ \mathtt{Loves}(\mathtt{y}, \mathtt{x})$). Further, unlike in Markov logic, first-order formulas in MEBNs must be hard (and consistent). Laskey and Costa do not specify a language for specifying conditional distributions; they simply assume that a terminating algorithm for computing them exists. Thus the question of what infinite distributions can be specified by MEBNs remains open.

6 Conclusion

In this paper, we extended the semantics of Markov logic to infinite domains using the theory of Gibbs measures. We gave sufficient conditions for the existence and uniqueness of a measure consistent with the local potentials defined by an MLN. We also described the structure of the set of consistent measures when it is not a singleton, and showed how the problem of satisfiability can be cast in terms of MLN measures. Directions for future work include designing lifted inference and learning algorithms for infinite MLNs, deriving alternative conditions for existence and uniqueness, analyzing the structure of consistent measure sets in more detail, extending the theory to non-Herbrand interpretations and recursive random fields (Lowd & Domingos, 2007), and studying interesting special cases of infinite MLNs.

References

Bacchus, F. (1990). *Representing and reasoning with probabilistic knowledge*. Cambridge, MA: MIT Press.

Bacchus, F., Grove, A. J., Halpern, J. Y., & Koller, D. (1996). From statistical knowledge bases to degrees of belief. *Artificial Intelligence*, *87*, 75–143.

Billingsley, P. (1995). *Probability and measure*. New York: John Wiley and Sons.

Booth, T., & Thompson, R. (1973). Applying probability measures to abstract languages. *IEEE Transactions on Computers*, *C-22*, 442–450.

Chi, Z. (1999). Statistical properties of probabilistic context-free grammars. *Computational Linguistics*, *25*, 131–160.

Dietterich, T., Getoor, L., & Murphy, K. (Eds.). (2003). *Proceedings of the ICML-2004 Workshop on Statistical Relational Learning and its Connections to Other Fields*. Banff, Canada: IMLS.

Domingos, P., Kok, S., Poon, H., Richardson, M., & Singla, P. (2006). Unifying logical and statistical AI. *Proceedings of the Twenty-First National Conference on Artificial Intelligence* (pp. 2–7). Boston, MA: AAAI Press.

Friedman, N., Getoor, L., Koller, D., & Pfeffer, A. (1999). Learning probabilistic relational models. *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence* (pp. 1300–1307). Stockholm, Sweden: Morgan Kaufmann.

Genesereth, M. R., & Nilsson, N. J. (1987). *Logical foundations of artificial intelligence*. San Mateo, CA: Morgan Kaufmann.

Georgii, H. (1988). *Gibbs measures and phase transitions*. Berlin: Walter de Gruyter.

Halpern, J. (1990). An analysis of first-order logics of probability. *Artificial Intelligence*, *46*, 311–350.

Jaeger, M. (1998). Reasoning about infinite random structures with relational Bayesian networks. *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning*. Trento, Italy: Morgan Kaufmann.

Kersting, K., & De Raedt, L. (2001). Towards combining inductive logic programming with Bayesian networks. *Proceedings of the Eleventh International Conference on Inductive Logic Programming* (pp. 118–131). Strasbourg, France: Springer.

Kok, S., Sumner, M., Richardson, M., Singla, P., Poon, H., & Domingos, P. (2006). *The Alchemy system for statistical relational AI* (Technical Report). Department of Computer Science and Engineering, University of Washington, Seattle, WA. http://alchemy.cs.washington.edu.

Laskey, K., & Costa, P. (2005). Of starships and Klingons: Bayesian logic for 23rd century. *Proceedings of the Twenty First Conference on Uncertainty in Artificial Intelligence* (pp. 346–353). Edinburgh, Scotland: UAI Press.

Lowd, D., & Domingos, P. (2007). Recursive random fields. *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*. Hyderabad, India: AAAI Press.

Milch, B., Marthi, B., Sontag, D., Russell, S., & Ong, D. L. (2005). Approximate inference for infinite contingent Bayesian networks. *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*. Barbados.

Muggleton, S. (1996). Stochastic logic programs. In L. De Raedt (Ed.), *Advances in inductive logic programming*, 254–264. Amsterdam, Netherlands: IOS Press.

Nilsson, N. (1986). Probabilistic logic. *Artificial Intelligence*, *28*, 71–87.

Pfeffer, A. (2000). *Probabilistic reasonsing for complex systems*. Doctoral dissertation, Department of Computer Science, Stanford University, Stanford, CA.

Pfeffer, A., & Koller, D. (2000). Semantics and inference for recursive probability models. *Proceedings of the Seventeenth National Conference on Artificial Intelligence* (pp. 538–544). Austin, TX: AAAI Press.

Poole, D. (1997). The independent choice logic for modelling multiple agents under uncertainty. *Artificial Intelligence*, *94*, 7–56.

Richardson, M., & Domingos, P. (2006). Markov logic networks. *Machine Learning*, *62*, 107–136.

Sato, T., & Kameya, Y. (1997). PRISM: A symbolic-statistical modeling language. *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence* (pp. 1330–1335). Nagoya, Japan: Morgan Kaufmann.

Taskar, B., Abbeel, P., & Koller, D. (2002). Discriminative probabilistic models for relational data. *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence* (pp. 485–492). Edmonton, Canada: Morgan Kaufmann.

Wellman, M., Breese, J. S., & Goldman, R. P. (1992). From knowledge bases to decision models. *Knowledge Engineering Review*, *7*.