

Linear Programs for Measuring Inconsistency in Probabilistic Logics

Nico Potyka

Department of Computer Science
FernUniversität in Hagen,
Germany

Abstract

Inconsistency measures help analyzing contradictory knowledge bases and resolving inconsistencies. In recent years several measures with desirable properties have been proposed, but often these measures correspond to combinatorial or non-convex optimization problems that are hard to solve in practice. In this paper, I study a new family of inconsistency measures for probabilistic knowledge bases. All members satisfy many desirable properties and can be computed by means of convex optimization techniques. For two members, I present linear programs whose computation is barely harder than a probabilistic satisfiability test.

1 Introduction

A central problem in the design of expert systems is to assure consistency of the knowledge base. The represented knowledge may originate from several experts or different sources of statistical data, which might be incompatible with each other. Even for a single expert designing a consistent knowledge base can be difficult.

This work focuses on probabilistic knowledge bases. Some heuristic approaches have been developed to resolve inconsistencies by adapting probabilities or deleting parts of the knowledge base (Rödger and Xu 1999; Finthammer, Kern-Isberner, and Ritterskamp 2007), but such methods do not help understanding possible causes for contradictions and lack a well-founded methodology. Inconsistency measures evaluate the degree of inconsistency (Knight 2002; Grant and Hunter 2008; Thimm 2013) and in this way can complement and guide heuristic approaches; however, many existing measures rely on combinatorial or non-convex optimization problems that can be hard to solve in practice.

In this paper, I study a new family of inconsistency measures that correspond to convex optimization problems. As opposed to non-convex problems, they lack (non-global) local minima and therefore are significantly easier to compute. Two members can be computed by particularly efficient linear programming techniques. The measures are based on the fact that a probabilistic knowledge base is satisfiable if and only if a corresponding system of linear equations can be solved. If the equation system is unsolvable, we can search

for a vector that violates the system in a minimal way. The overall violation can be measured with respect to some norm like the euclidean norm. If the violation value is zero, the knowledge base is consistent. Otherwise, the value increases continuously with the degree of violation. This is the basic idea of the family of *minimal violation measures*.

Inconsistency Measures for Probabilistic Logics

In recent years various approaches have been developed to measure the degree to which knowledge bases are inconsistent, see, e.g., (Knight 2002; Grant and Hunter 2008; Doder et al. 2010). A short overview of measures for classical logics can be found in (Hunter and Konieczny 2006). Extensions and new measures for probabilistic logics are summarized in (Thimm 2013). I quickly sketch some examples that will be referred to in the following.

The easiest measure one can possibly think of is the *drastic measure* (Hunter and Konieczny 2006; Thimm 2013) that yields 0 if the knowledge base is consistent and 1 otherwise.

A class of measures that is interesting for both classical and probabilistic knowledge bases relies on *minimal inconsistent sets* (Hunter and Konieczny 2008; Thimm 2013). A subset of a knowledge base is called a minimal inconsistent set, if it is inconsistent and removing a single rule renders the set consistent. Hence, the contained rules are in conflict with each other and the conflict can be resolved by removing a single rule. Interesting measures can be defined by counting the minimal inconsistent sets, or by combining the numbers of minimal inconsistent sets that contain certain rules (Hunter and Konieczny 2008; Thimm 2013).

A particularly interesting class of measures derives from the idea of *minimally changing* the knowledge base to a consistent one. In (Thimm 2009), the probabilities of the rules are changed and an inconsistency measure is defined by the minimal sum of all absolute changes. In (Picado-Muñoz 2011) the probabilities are relaxed to probability intervals. The corresponding inconsistency measure is defined by arranging the size of these intervals in a vector and measuring the vector's size with respect to a norm. Both measures are particularly interesting because they yield an *optimal repair* for a knowledge base, that is, a consistent knowledge base that is obtained from an inconsistent one by minimal changes.

Computational Aspects: The drastic measure corresponds to a satisfiability test and can be regarded as the measure that is most easily computable, because each reasonable measure should distinguish consistent from inconsistent sets. Deciding whether a probabilistic knowledge base is inconsistent is NP-hard in general (Georgakopoulos, Kavvadias, and Papadimitriou 1988). Still, for knowledge bases containing more than one hundred propositional variables and several hundred rules satisfiability can be tested by linear programming techniques in reasonable time (Jaumard, Hansen, and Poggi 1991; Hansen and Perron 2008).

Measures based on minimal inconsistent sets correspond to combinatorial optimization problems. There has been some progress in implementing such measures efficiently for classical logics (Konieczny and Roussel 2013), but the key idea is that a classical knowledge base can be regarded as a single formula in conjunctive normal form. This, however, does not apply to probabilistic knowledge bases. As the number of subsets of a knowledge base is exponential in the size of the knowledge base, the vital problems are efficiently enumerating candidates for minimal inconsistent sets and testing their satisfiability.

Minimal change measures like in (Thimm 2009) and (Picado-Muñoz 2011) correspond to non-convex optimization problems that suffer from (non-global) local optima. Therefore, computing these measures can be hard in practice.

Overview of the Paper

The logical framework and some notations are explained in Section 2. In Section 3, I introduce the family of minimal violation measures and demonstrate in some examples how they differ from other measures and how they can be used to analyze and to resolve inconsistencies. In Section 4, I show that minimal violation measures satisfy several desirable properties that have been considered in the literature (Knight 2003; Hunter and Konieczny 2006; Thimm 2013). Section 5 deals with computational issues. I present linear programs for two minimal violation measures and compare their performance to other measures using prototypical implementations. Subsequently, I summarize some empirical estimates for the performance of typical optimization algorithms (Boyd and Vandenberghe 2004; Matousek and Gärtner 2007) and use them to provide some rough estimates for the performance of different measures when applying general purpose algorithms.

2 Preliminaries

Consider a propositional language $\mathcal{L}_\Sigma$ built up over a finite set of propositional variables Σ using logical connectives like conjunction and disjunction in the usual way. For formulas $\mathcal{F}, \mathcal{G} \in \mathcal{L}_\Sigma$, negation $\neg \mathcal{F}$ is abbreviated by an overbar $\overline{\mathcal{F}}$ and conjunction $\mathcal{F} \wedge \mathcal{G}$ by juxtaposition $\mathcal{F}\mathcal{G}$. A *possible world* is a classical logical interpretation $\omega : \Sigma \rightarrow \{0, 1\}$ assigning a truth value to each propositional variable. Let Ω denote the set of all possible worlds. An atom a is satisfied by ω iff $\omega(a) = 1$. The definition is extended to complex formulas in the usual way. The *classical models*

$\text{Mod}(\mathcal{G})$ of a formula $\mathcal{G}$ are the possible worlds $\omega \in \Omega$ that satisfy $\mathcal{G}$. $\mathcal{F}, \mathcal{G} \in \mathcal{L}_\Sigma$ are called *equivalent*, $\mathcal{F} \equiv \mathcal{G}$, iff $\text{Mod}(\mathcal{F}) = \text{Mod}(\mathcal{G})$.

Consider the probabilistic conditional language $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma) := \{(\mathcal{G} | \mathcal{F})[\rho] \mid \mathcal{G}, \mathcal{F} \in \mathcal{L}_\Sigma, \rho \in [0, 1]\}$ over Σ (Lukasiewicz 1999). Intuitively, a conditional $(\mathcal{G} | \mathcal{F})[\rho]$ expresses that our belief in $\mathcal{G}$ given that $\mathcal{F}$ holds is ρ . Unconditioned probabilistic formulas are obtained for the special case $(\mathcal{G} | \top)[\rho]$, where $\top$ denotes a tautological formula. $(\mathcal{G} | \top)[\rho]$ is abbreviated by $(\mathcal{G})[\rho]$.

Probabilistic interpretations $\mathcal{P} : \Omega \rightarrow [0, 1]$ are probability functions over possible worlds. For a formula $\mathcal{G} \in \mathcal{L}_\Sigma$ let $\mathcal{P}(\mathcal{G}) = \sum_{\omega \in \text{Mod}(\mathcal{G})} \mathcal{P}(\omega)$. A probabilistic interpretation $\mathcal{P}$ satisfies a conditional $(\mathcal{G} | \mathcal{F})[\rho]$ iff $\mathcal{P}(\mathcal{G}\mathcal{F}) = \rho \cdot \mathcal{P}(\mathcal{F})$. If $\mathcal{P}(\mathcal{F}) > 0$ this definition corresponds to the definition of conditional probability $\mathcal{P}(\mathcal{G} | \mathcal{F}) = \frac{\mathcal{P}(\mathcal{G}\mathcal{F})}{\mathcal{P}(\mathcal{F})} = \rho$. As often done (Paris and Vencovská 1989; Lukasiewicz 1999; Thimm 2013), I do not demand $\mathcal{P}(\mathcal{F}) > 0$.

A *probabilistic knowledge base* $\mathcal{R} \subset (\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$ is a finite set of conditionals. A probabilistic interpretation $\mathcal{P}$ satisfies $\mathcal{R}$ iff it satisfies each $c \in \mathcal{R}$. Let $\text{Mod}(\mathcal{R})$ denote the set of probabilistic models of $\mathcal{R}$, i.e., the set of probabilistic interpretations over Ω that satisfy $\mathcal{R}$.

Two conditionals $c_1, c_2 \in (\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$ are called *equivalent*, $c_1 \equiv c_2$, iff $\text{Mod}(c_1) = \text{Mod}(c_2)$. Following (Thimm 2013), two knowledge bases $\mathcal{R}_1, \mathcal{R}_2$ over $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$ are called (*semi-extensionally*) *equivalent*, $\mathcal{R}_1 \equiv \mathcal{R}_2$, iff there is a bijection $B : \mathcal{R}_1 \rightarrow \mathcal{R}_2$ such that $c_1 \equiv B(c_1)$ for each $c_1 \in \mathcal{R}_1$. That is, both knowledge bases contain the same number of conditionals and these conditionals differ only in their syntactical representation. If both $\mathcal{G}\mathcal{F}$ and $\overline{\mathcal{G}}\mathcal{F}$ are satisfiable, $(\mathcal{G} | \mathcal{F})[\rho]$ is called a *normal conditional*. The following lemma states an intermediate result that was obtained in (Thimm 2013) in the proof of Theorem 2, (4).

Lemma 2.1. (Thimm 2013) *Let $c_1, c_2 \in \mathcal{R}$ be normal, $c_i = (\mathcal{G}_i | \mathcal{F}_i)[\rho_i]$. If $c_1 \equiv c_2$, then $\mathcal{F}_1 \equiv \mathcal{F}_2$ and either*

- $\mathcal{G}_1\mathcal{F}_1 \equiv \mathcal{G}_2\mathcal{F}_2$ and $\rho_1 = \rho_2$, or
- $\mathcal{G}_1\mathcal{F}_1 \equiv \overline{\mathcal{G}_2}\mathcal{F}_2$ and $\rho_1 = 1 - \rho_2$.

As done in (Thimm 2013), I will assume that knowledge bases contain only normal conditionals.

If $\text{Mod}(\mathcal{R}) \neq \emptyset$, $\mathcal{R}$ is called *consistent*. Given a consistent knowledge base, we can draw probabilistic conclusions in different ways. For example, we can consider the whole set $\text{Mod}(\mathcal{R})$ and compute probability intervals for arbitrary formulas, see, e.g., (Nilsson 1986; Jaumard, Hansen, and Poggi 1991). Another possible way is to select a single best model from $\text{Mod}(\mathcal{R})$ with respect to some evaluation function and to use it to compute point probabilities for arbitrary (conditional) formulas, see, e.g., (Paris and Vencovská 1989; Kern-Isberner 2001).

An inconsistency measure assigns a nonnegative inconsistency value to each knowledge base.

Definition 2.2. *An inconsistency measure Inc is a function $\text{Inc} : 2^{(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)} \rightarrow \mathbb{R}_{\geq 0}$ such that $\text{Inc}(\mathcal{R}) = 0$ if and only if $\mathcal{R}$ is consistent.*

Several interesting properties have been considered in the literature to motivate and to compare different inconsistency measures, see, e.g., (Knight 2002; Hunter and Konieczny 2006; Picado-Muñoz 2011). An overview of inconsistency measures for probabilistic logics and a comparison of some of their properties can be found in (Thimm 2013).

Matrix representation of knowledge bases

Sometimes it is more convenient to represent the conditional satisfaction constraints in matrix notation (Nilsson 1986; Jaumard, Hansen, and Poggi 1991). Suppose that the possible worlds in Ω are ordered in an arbitrary but fixed sequence $\omega_1, \dots, \omega_n$, $n = |\Omega|$, so that each probabilistic interpretation $\mathcal{P}$ can be identified with a column vector $\vec{\pi} = (\mathcal{P}(\omega_1) \dots \mathcal{P}(\omega_n))^T$, where the symbol T denotes transposition of vectors. The set of all probabilistic interpretations of $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$ corresponds to the set $\Pi_n = \{\vec{\pi} \in \mathbb{R}^n \mid \vec{\pi} \geq \vec{0} \text{ and } \sum_{i=1}^n \pi_i = 1\}$, where $\geq$ denotes componentwise inequality.

To compare vectors, we can consider different norms on $\mathbb{R}^n$. A p -norm is a function $\|\cdot\|_p : \mathbb{R}^n \rightarrow \mathbb{R}$, which is defined by $\|\vec{\pi}\|_p = \sqrt[p]{\sum_{i=1}^n |\pi_i|^p}$. The best known special cases are the 1-norm $\|\vec{\pi}\|_1 = \sum_{i=1}^n |\pi_i|$, the euclidean norm $\|\vec{\pi}\|_2 = \sqrt{\sum_{i=1}^n \pi_i^2}$ and the limit for $p \rightarrow \infty$, the maximum norm $\|\vec{\pi}\|_\infty = \max\{|\pi_i| \mid 1 \leq i \leq n\}$.

Recall that a probabilistic model of c satisfies the equation $\mathcal{P}(\mathcal{G}\mathcal{F}) = \rho \cdot \mathcal{P}(\mathcal{F})$. Subtracting $\rho \cdot \mathcal{P}(\mathcal{F})$ yields

$$\begin{aligned} 0 &= \mathcal{P}(\mathcal{G}\mathcal{F}) - \rho \cdot \mathcal{P}(\mathcal{F}) \\ &= \mathcal{P}(\mathcal{G}\mathcal{F}) - \rho \cdot (\mathcal{P}(\mathcal{G}\mathcal{F}) + \mathcal{P}(\bar{\mathcal{G}}\mathcal{F})) \\ &= (1 - \rho) \cdot \mathcal{P}(\mathcal{G}\mathcal{F}) - \rho \cdot \mathcal{P}(\bar{\mathcal{G}}\mathcal{F}) \\ &= \sum_{\omega \in \text{Mod}(\mathcal{G}\mathcal{F})} \mathcal{P}(\omega) \cdot (1 - \rho) - \sum_{\omega \in \text{Mod}(\bar{\mathcal{G}}\mathcal{F})} \mathcal{P}(\omega) \cdot \rho \\ &= \sum_{\omega \in \Omega} \mathcal{P}(\omega) \cdot (1_{\{\mathcal{G}\mathcal{F}\}}(\omega) \cdot (1 - \rho) - 1_{\{\bar{\mathcal{G}}\mathcal{F}\}}(\omega) \cdot \rho), \end{aligned}$$

where for a formula F the indicator function $1_{\{F\}}(\omega)$ maps to 1 iff $\omega \models_{\mathcal{L}_\Sigma} F$ and to 0 otherwise. In vector notation this equation becomes

$$\vec{a}_c^T \vec{\pi} = 0,$$

where $\vec{a}_c^T$ is the transpose of the vector

$$(1_{\{\mathcal{G}\mathcal{F}\}}(\omega_j) \cdot (1 - \rho) - 1_{\{\bar{\mathcal{G}}\mathcal{F}\}}(\omega_j) \cdot \rho)_{1 \leq j \leq n}.$$

That is, the j -th component of $\vec{a}^T$ is either $1 - \rho$, $-\rho$ or 0 dependent on whether the conditional is verified ($1_{\{\mathcal{G}\mathcal{F}\}}(\omega_j) = 1$), falsified ($1_{\{\bar{\mathcal{G}}\mathcal{F}\}}(\omega_j) = 1$) or not applicable ($\omega_j \models_{\mathcal{L}_\Sigma} \bar{\mathcal{F}}$) in the j -th world.

Given a knowledge base $\mathcal{R}$, suppose that the conditionals in $\mathcal{R}$ are ordered in an arbitrary but fixed sequence $c_1, \dots, c_m$, $m = |\mathcal{R}|$, so that we can identify the i -th conditional with a row vector $\vec{a}_i^T$. Then we can associate $\mathcal{R}$ with the $(m \times n)$ -matrix

$$A_{\mathcal{R}} = \begin{pmatrix} \vec{a}_1^T \\ \vdots \\ \vec{a}_m^T \end{pmatrix}.$$

Using the introduced notations, the question whether $\mathcal{R}$ is consistent becomes the question whether the linear equation system $A_{\mathcal{R}}x = 0$ can be solved by a probability vector $\vec{\pi} \in \Pi_n$ (Nilsson 1986; Jaumard, Hansen, and Poggi 1991).

3 Minimal Violation Measures

Given a knowledge base $\mathcal{R}$ over $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$, the corresponding matrix $A_{\mathcal{R}}$ and some p -norm $\|\cdot\|_p$ as explained in Section 2, we consider the following optimization problem:

$$\begin{aligned} \min_{\vec{\pi} \in \mathbb{R}^n} \quad & \|A_{\mathcal{R}}\vec{\pi}\|_p \\ \text{subject to} \quad & \sum_{i=1}^n \pi_i = 1 \\ & \vec{\pi} \geq 0. \end{aligned} \quad (1)$$

The constraints just guarantee that the solution is a probabilistic interpretation in Π_n . The objective $\|A_{\mathcal{R}}\vec{\pi}\|_p$ measures to which degree $\vec{\pi}$ violates $\mathcal{R}$ with respect to the p -norm. The choice of p depends on the application. For $p = 1$ a lower total violation might be obtained but the violation can be rather extreme. As p grows, higher violations are penalized more heavily. The following example illustrates this behavior for the extreme cases $p = 1$ and $p = \infty$.

Example 3.1. Consider the vectors $v_1 = (\frac{1}{4} \ \frac{1}{4} \ \frac{1}{4} \ \frac{1}{4})^T$, $v_2 = (0 \ 0 \ 0 \ \frac{1}{2})^T$ in $\mathbb{R}^4$. v_1 corresponds to a violation of $\frac{1}{4}$ for four conditionals. In v_2 no conditional but the last one is violated. We have $\|v_1\|_1 = 1$, $\|v_2\|_1 = \frac{1}{2}$. Hence v_2 is preferable to v_1 with respect to $\|\cdot\|_1$, because the overall error is lower. In contrast, $\|v_1\|_\infty = \frac{1}{4}$, $\|v_2\|_\infty = \frac{1}{2}$. Hence v_1 is preferable to v_2 with respect to $\|\cdot\|_\infty$, because the highest error is lower. Which behavior is more desirable, depends on the specific application.

The minimal violation inconsistency measures are defined by means of optimization problem (1). A vector $\vec{\pi} \in \Pi_n$ minimizing optimization problem (1) will be called *minimal violating* in the following. If $\mathcal{R}$ is consistent, the minimal violating $\vec{\pi}$ are the models of $\mathcal{R}$. For inconsistent $\mathcal{R}$, they can be regarded to be as close to a model as possible.

Definition 3.2. For $p \in \mathbb{N} \cup \{\infty\}$ the minimal violation p -inconsistency measure $\text{Inc}_{\Pi}^p : 2^{(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)} \rightarrow \mathbb{R}$ is defined by $\text{Inc}_{\Pi}^p(\mathcal{R}) = \epsilon$, where ϵ is the solution of optimization problem (1).

Proposition 3.3. Inc_{Π}^p is an inconsistency measure.

Proof. To begin with, we convince ourselves that Inc_{Π}^p is well-defined. That is, there always exists a unique minimum for optimization problem (1). The objective is the composition of a p -norm and the linear function induced by the constraint matrix $A_{\mathcal{R}}$ and therefore continuous and convex. The feasible region of optimization problem (1) is the convex and compact probability simplex. As each continuous function on a compact set takes on its minimum, the minimum exists. It is also unique, because the minimum for each convex function on a convex set is unique.

It remains to show that Inc_{Π}^p is an inconsistency measure as defined in Definition 2.2. By definition of optimization

problem (1), it holds $\text{Inc}_{\Pi}^p(\mathcal{R}) \geq 0$. $\mathcal{R}$ is consistent if and only if there is a $\vec{\pi} \in \Pi_n$ such that $A_{\mathcal{R}}\vec{\pi} = 0$. Hence, by definiteness of norms, it follows $\text{Inc}_{\Pi}^p(\mathcal{R}) = 0$ if and only if $\mathcal{R}$ is consistent. $\square$

Before investigating general properties of Inc_{Π}^p , we consider some simple examples and explore differences between Inc_{Π}^p and the drastic measure (Hunter and Konieczny 2006; Thimm 2013), measures based on minimal inconsistent sets (Hunter and Konieczny 2008; Thimm 2013) as well as minimal change measures (Thimm 2009; Picado-Muñoz 2011) that have been explained in Section 1.

Example 3.4. Consider $\Sigma = \{A\}$ and the knowledge base $\mathcal{R}_{\rho} = \{(A)[\rho], (A)[1 - \rho]\}$ over Σ for $0 \leq \rho \leq 0.5$. For instance, $\mathcal{R}_{0.5} = \{(A)[0.5], (A)[0.5]\} = \{(A)[0.5]\}$ is consistent and $\mathcal{R}_0 = \{(A)[0], (A)[1]\}$ is inconsistent. Intuitively, the degree of inconsistency of $\mathcal{R}_{\rho}$ should increase as ρ decreases from 0.5 to 0. The following table shows some inconsistency values of $\mathcal{R}_{\rho}$ with respect to Inc_{Π}^p .

	$\mathcal{R}_{0.5}$	$\mathcal{R}_{0.49}$	$\mathcal{R}_{0.4}$	$\mathcal{R}_{0.2}$	$\mathcal{R}_0$
Inc_{Π}^1	0	0.02	0.2	0.6	1
$\text{Inc}_{\Pi}^{\infty}$	0	0.01	0.1	0.3	0.5

Minimal change measures also capture this intuitive increase of inconsistency. In contrast, the drastic measure and inconsistency measures based on minimal inconsistent sets cannot distinguish the knowledge bases $\mathcal{R}_{\rho}$ for $0 \leq \rho < 0.5$, because all these $\mathcal{R}_{\rho}$ are inconsistent and in each case the only minimal inconsistent set is $\mathcal{R}_{\rho}$ itself.

Example 3.5. Consider $\Sigma = \{A, B\}$ and the knowledge base $\mathcal{R} = \{(A)[0.8], (B)[0.6], (B | A)[0.9]\}$. We find

$$\text{Inc}_{\Pi}^1(\mathcal{R}) \approx 0.120 \text{ and } \text{Inc}_{\Pi}^{\infty}(\mathcal{R}) \approx 0.041.$$

Intuitively, we can conjecture that decreasing the probability of A , increasing the probability of B or decreasing the probability of B given A might resolve the contradiction. We start by decreasing the probability of A and obtain for $\mathcal{R}_1 = \{(A)[0.75], (B)[0.6], (B | A)[0.9]\}$

$$\text{Inc}_{\Pi}^1(\mathcal{R}_1) \approx 0.075 \text{ and } \text{Inc}_{\Pi}^{\infty}(\mathcal{R}_1) \approx 0.025.$$

As suspected, the degree of inconsistency decreased. Next, we increase the probability of B and obtain for $\mathcal{R}_2 = \{(A)[0.75], (B)[0.65], (B | A)[0.9]\}$

$$\text{Inc}_{\Pi}^1(\mathcal{R}_2) \approx 0.024 \text{ and } \text{Inc}_{\Pi}^{\infty}(\mathcal{R}_2) \approx 0.008.$$

Again, the degree of inconsistency decreased significantly. Now we decrease the probability of B given A and obtain for $\mathcal{R}_3 = \{(A)[0.75], (B)[0.65], (B | A)[0.85]\}$

$$\text{Inc}_{\Pi}^1(\mathcal{R}_3) = 0 \text{ and } \text{Inc}_{\Pi}^{\infty}(\mathcal{R}_3) = 0.$$

Hence $\mathcal{R}_3$ yields a repair for $\mathcal{R}$.

If we had no idea, which conditionals might have caused the inconsistencies and in which way, we could try to adjust probabilities arbitrarily. For example, if we increased the

probability of A in the first step instead of decreasing it, we would obtain with $\mathcal{R}_0 = \{(A)[0.85], (B)[0.6], (B | A)[0.9]\}$

$$\text{Inc}_{\Pi}^1(\mathcal{R}_0) \approx 0.165 \text{ and } \text{Inc}_{\Pi}^{\infty}(\mathcal{R}_0) \approx 0.057.$$

So the inconsistency values increased compared to our initial knowledge base $\mathcal{R}$ and we can see that this is not a reasonable adjustment. In the same way, we can learn that decreasing the probability of B or increasing the probability of B given A aggravates the conflict.

In contrast, the drastic measure provides no interesting information to resolve such inconsistencies. Measures based on minimal inconsistent sets are more helpful. $\mathcal{R}$ is the only minimal inconsistent set, so we could repair $\mathcal{R}$ by deleting a single conditional. Furthermore, we learn that all conditionals are in conflict with each other, so each adjustment might be reasonable. However, $\mathcal{R}$, $\mathcal{R}_0$ and $\mathcal{R}_1$ cannot be distinguished by measures based on minimal inconsistent sets, because in each case the knowledge base itself is the only minimal inconsistent subset. Therefore, they cannot provide information about reasonable adjustments, unless a conflict has been resolved by the adjustment coincidentally.

The best choice in this scenario is admittedly a minimal change measure, because it yields a best repair without further ado. A best repair (there can be several) computed by the measure from (Thimm 2009) is $\mathcal{R}_{\text{opt}} = \{(A)[0.752], (B)[0.648], (B | A)[0.86]\}$. As we will see later on, minimal violation measures become superior to minimal change measures for larger knowledge bases because of computational advantages.

4 Properties of Minimal Violation Measures

We now turn to some properties that have been considered interesting for inconsistency measures in the literature (Knight 2003; Hunter and Konieczny 2006; Thimm 2013).

To give an intuitive overview, Inc_{Π}^p is language and representation invariant. That is, the degree of inconsistency of a knowledge base determined by Inc_{Π}^p is independent of the number of propositional variables in the underlying language and the syntactical representation of the conditionals. Inc_{Π}^p is monotone and for $p = 1$ super-additive. The former means that adding new conditionals can only increase the degree of inconsistency. The latter means that if we combine two knowledge bases, the inconsistency value of the combined knowledge base is at least as large as the sum of both single inconsistency values. Furthermore, Inc_{Π}^p is weakly independent. That is, if we add consistent knowledge about propositional variables that are not contained in our current knowledge base, the inconsistency value does not increase. The term *weak* stems from the fact that stronger independence properties for inconsistent sets have been considered in the literature (Hunter and Konieczny 2006; Thimm 2013). Finally, Inc_{Π}^p is continuous. That is, small adjustments in the probabilities of the knowledge base change the corresponding inconsistency value only slightly. Some other properties have been considered in (Knight 2003; Hunter and Konieczny 2006; Thimm 2013), but in my opinion the stated ones are among the most intuitive for probabilistic knowledge bases and I leave it for future work to

answer whether minimal violation measures satisfy or violate further properties.

We need some additional notation in this section. For a conditional $c \in (\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$ let $\text{At}(c)$ denote the atoms from Σ that appear in c and for a knowledge base $\mathcal{R}$ let $\text{At}(\mathcal{R})$ denote the atoms appearing in $\mathcal{R}$. Note that we can consider the same knowledge base $\mathcal{R}$ with respect to different languages $(\Sigma' | \Sigma')$ as long as $\text{At}(\mathcal{R}) \subseteq \Sigma'$. We will see that the size of Σ' does not influence the inconsistency value with respect to Inc_Π^p , but to prove this fact, we have to make a distinction between the same knowledge base in different languages. For a knowledge base $\mathcal{R}$ that is defined with respect to a language $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$, I call $\mathcal{R}$ *restricted to* $(\Sigma' | \Sigma')$ if $\Sigma' \subseteq \Sigma$ and I call $\mathcal{R}$ *lifted to* $(\Sigma' | \Sigma')$ if $\Sigma \subseteq \Sigma'$.

We first consider language invariance, which means that changing our language does not change the inconsistency value of a knowledge base (Knight 2003).

Lemma 4.1. *Let V be a new propositional variable not contained in Σ and let $\Sigma' = \Sigma \cup \{V\}$. Let $\mathcal{R}$ be a knowledge base over $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$ and let $\mathcal{R}'$ denote $\mathcal{R}$ lifted to $(\Sigma' | \Sigma')$. Then $\text{Inc}_\Pi^p(\mathcal{R}) = \text{Inc}_\Pi^p(\mathcal{R}')$.*

Proof. Let $\omega_1, \dots, \omega_n$ be the ordered possible worlds over Σ . Let $\omega'_1, \dots, \omega'_n, \omega'_{n+1}, \dots, \omega'_{2n}$ denote the possible worlds over Σ' , such that ω'_i, ω'_{i+n} correspond to ω_i for the variables in Σ , $\omega'_1, \dots, \omega'_n$ falsify V , and $\omega'_{n+1}, \dots, \omega'_{2n}$ satisfy V . Note that $A_{\mathcal{R}'} = (A_{\mathcal{R}} \ A_{\mathcal{R}})$ by construction, because $\mathcal{R}'$ corresponds to $\mathcal{R}$ and does not contain V .

Let $\text{Inc}_\Pi^p(\mathcal{R}) = \delta$ and let $\vec{\pi}^*$ be minimal violating, i.e., $\|A_{\mathcal{R}} \vec{\pi}^*\|_p = \delta$. Then $\begin{pmatrix} \vec{\pi}^* \\ \vec{0} \end{pmatrix} \in \mathbb{R}^{2n}$ corresponds to a probability function over the interpretations of Σ' and

$$\begin{aligned} \text{Inc}_\Pi^p(\mathcal{R}') &\leq \| (A_{\mathcal{R}} \ A_{\mathcal{R}}) \begin{pmatrix} \vec{\pi}^* \\ \vec{0} \end{pmatrix} \|_p = \|A_{\mathcal{R}} \vec{\pi}^*\|_p \\ &= \text{Inc}_\Pi^p(\mathcal{R}). \end{aligned}$$

Conversely, if $\text{Inc}_\Pi^p(\mathcal{R}') = \delta'$ and $\begin{pmatrix} \vec{\pi}_1^* \\ \vec{\pi}_2^* \end{pmatrix} \in \mathbb{R}^{2n}$ is minimal violating, then $(\vec{\pi}_1^* + \vec{\pi}_2^*) \in \mathbb{R}^n$ corresponds to a probability function over the interpretations of Σ and

$$\begin{aligned} \text{Inc}_\Pi^p(\mathcal{R}) &\leq \|A_{\mathcal{R}} (\vec{\pi}_1^* + \vec{\pi}_2^*)\|_p = \|A_{\mathcal{R}} \vec{\pi}_1^* + A_{\mathcal{R}} \vec{\pi}_2^*\|_p \\ &= \| (A_{\mathcal{R}} \ A_{\mathcal{R}}) \begin{pmatrix} \vec{\pi}_1^* \\ \vec{\pi}_2^* \end{pmatrix} \|_p = \text{Inc}_\Pi^p(\mathcal{R}'). \end{aligned}$$

Hence, $\text{Inc}_\Pi^p(\mathcal{R}') \leq \text{Inc}_\Pi^p(\mathcal{R}) \leq \text{Inc}_\Pi^p(\mathcal{R}')$, that is, both inconsistency values have to be equal. $\square$

Proposition 4.2 (Language invariance). *Let $\mathcal{R}$ be a knowledge base over $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$ and let $\text{At}(\mathcal{R}) \subseteq \Sigma'$. Let $\mathcal{R}'$ denote $\mathcal{R}$ over $(\Sigma' | \Sigma')$. Then $\text{Inc}_\Pi^p(\mathcal{R}) = \text{Inc}_\Pi^p(\mathcal{R}')$.*

Proof. For $\Sigma \subseteq \Sigma'$, the claim follows from Lemma 4.1 by induction over the variables in $\Sigma' \setminus \Sigma$.

For the general case, consider $\Sigma'' = \Sigma \cap \Sigma'$ and let $\mathcal{R}'' \subseteq (\Sigma'' | \Sigma'')$ denote $\mathcal{R}$ restricted to $(\Sigma'' | \Sigma'')$. Then $\Sigma'' \subseteq \Sigma$ and $\Sigma'' \subseteq \Sigma'$ and therefore $\text{Inc}_\Pi^p(\mathcal{R}'') = \text{Inc}_\Pi^p(\mathcal{R}) = \text{Inc}_\Pi^p(\mathcal{R}')$. $\square$

A measure should also be invariant with respect to syntactical changes of the conditionals (Knight 2003). This property has also been called *irrelevance of syntax* (Thimm 2013).

Proposition 4.3 (Representation invariance). *Let $\mathcal{R}_1, \mathcal{R}_2$ be knowledge bases over $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$. If $\mathcal{R}_1 \equiv \mathcal{R}_2$, then $\text{Inc}_\Pi^p(\mathcal{R}_1) = \text{Inc}_\Pi^p(\mathcal{R}_2)$.*

Proof. As $\mathcal{R}_1 \equiv \mathcal{R}_2$, there is a bijection $B : \mathcal{R}_1 \rightarrow \mathcal{R}_2$ such that $c \equiv B(c)$ for each $c \in \mathcal{R}_1$. Consider an arbitrary conditional $c_1 \in \mathcal{R}_1$ such that $B(c_1) = c_2$. For $i = 1, 2$ let $c_i = (\mathcal{G}_i | \mathcal{F}_i)[\rho_i]$ and let $\vec{a}_i^T$ denote the corresponding constraint vectors. By Lemma 2.1, there are two possible cases. If $\mathcal{G}_1 \mathcal{F}_1 \equiv \mathcal{G}_2 \mathcal{F}_2$ and $\rho_1 = \rho_2$, then $\vec{a}_1^T = \vec{a}_2^T$. Otherwise, if $\mathcal{G}_1 \mathcal{F}_1 \equiv \overline{\mathcal{G}_2} \mathcal{F}_2$ and $\rho_1 = 1 - \rho_2$, then also $\mathcal{G}_2 \mathcal{F}_2 \equiv \overline{\mathcal{G}_1} \mathcal{F}_1$ because $\equiv$ is symmetric. Therefore, for the j -th component of the constraint vectors it holds

$$\begin{aligned} (\vec{a}_1^T)_j &= 1_{\{\mathcal{G}_1 \mathcal{F}_1\}}(\omega_j) \cdot (1 - \rho_1) - 1_{\{\overline{\mathcal{G}_1} \mathcal{F}_1\}}(\omega_j) \cdot \rho_1 \\ &= 1_{\{\overline{\mathcal{G}_2} \mathcal{F}_2\}}(\omega_j) \cdot \rho_2 - 1_{\{\mathcal{G}_2 \mathcal{F}_2\}}(\omega_j) \cdot (1 - \rho_2) \\ &= -(1_{\{\mathcal{G}_2 \mathcal{F}_2\}}(\omega_j) \cdot (1 - \rho_2) - 1_{\{\overline{\mathcal{G}_2} \mathcal{F}_2\}}(\omega_j) \cdot \rho_2) \\ &= -(\vec{a}_2^T)_j. \end{aligned}$$

Hence $\vec{a}_1^T = -\vec{a}_2^T$ and $|\vec{a}_1^T \vec{\pi}| = |-\vec{a}_2^T \vec{\pi}| = |\vec{a}_2^T \vec{\pi}|$ for all $\vec{\pi} \in \mathbb{R}^n$. As $c_1 \in \mathcal{R}_1$ was arbitrary, it follows $\text{Inc}_\Pi^p(\mathcal{R}_1) = \text{Inc}_\Pi^p(\mathcal{R}_2)$. $\square$

Monotonicity states that if we add additional conditionals to a knowledge base, the degree of inconsistency cannot decrease (Hunter and Konieczny 2006; Thimm 2013).

Proposition 4.4 (Monotonicity). *Let $\mathcal{R}, \mathcal{R}'$ be knowledge bases over $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$. If $\mathcal{R}' \subseteq \mathcal{R}$, then $\text{Inc}_\Pi^p(\mathcal{R}') \leq \text{Inc}_\Pi^p(\mathcal{R})$.*

Proof. Let $\{1, \dots, m\}$ denote the indices of the conditionals in $\mathcal{R}'$ and let $\{m+1, \dots, m+k\}$ denote the indices of the additional conditionals in $\mathcal{R} \setminus \mathcal{R}'$. Let $\vec{a}_i^T$ denote the corresponding row vectors and let $\vec{\pi}$ be minimal violating for $\mathcal{R}$. Then

$$\begin{aligned} \text{Inc}_\Pi^p(\mathcal{R}') &\leq \sqrt[p]{\sum_{i=1}^m |\vec{a}_i^T \vec{\pi}|^p} \leq \sqrt[p]{\sum_{i=1}^{m+k} |\vec{a}_i^T \vec{\pi}|^p} \\ &= \text{Inc}_\Pi^p(\mathcal{R}). \end{aligned}$$

$\square$

For $p = 1$, Inc_Π^p also satisfies a stronger property that has been called *super-additivity* in (Thimm 2013).

Proposition 4.5 (Super-Additivity). *Let $\mathcal{R}_1, \mathcal{R}_2$ be knowledge bases over $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$. If $\mathcal{R}_1 \cap \mathcal{R}_2 = \emptyset$, then $\text{Inc}_\Pi^1(\mathcal{R}_1 \cup \mathcal{R}_2) \geq \text{Inc}_\Pi^1(\mathcal{R}_1) + \text{Inc}_\Pi^1(\mathcal{R}_2)$.*

Proof. Let $\{1, \dots, m_1\}$ denote the indices of the conditionals in $\mathcal{R}_1$, and let $\{m_1+1, \dots, m_1+m_2\}$ denote the indices

of the conditionals in $\mathcal{R}_2$. Let $\vec{a}_i^\top$ denote the corresponding row vectors. Then

$$\begin{aligned} \text{Inc}_\Pi^1(\mathcal{R}_1 \cup \mathcal{R}_2) &= \min_{\vec{\pi} \in \mathbb{R}^n} \left(\sum_{i=1}^{m_1} |\vec{a}_i^\top \vec{\pi}| + \sum_{i=m_1+1}^{m_1+m_2} |\vec{a}_i^\top \vec{\pi}| \right) \\ &\geq \left(\min_{\vec{\pi} \in \mathbb{R}^n} \sum_{i=1}^{m_1} |\vec{a}_i^\top \vec{\pi}| \right) + \left(\min_{\vec{\pi} \in \mathbb{R}^n} \sum_{i=m_1+1}^{m_1+m_2} |\vec{a}_i^\top \vec{\pi}| \right) \\ &= \text{Inc}_\Pi^1(\mathcal{R}_1) + \text{Inc}_\Pi^1(\mathcal{R}_2). \end{aligned}$$

□

Remark 4.6. For $p > 1$ super-additivity can be violated.

If we add consistent knowledge about propositional variables that were not contained in our knowledge base before, we should expect that the knowledge base does not become more inconsistent. This intuition is captured by *weak independence* (Thimm 2013).

Proposition 4.7 (Weak Independence). *Let $\mathcal{R}_1, \mathcal{R}_2$ be knowledge bases over $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$. If $\mathcal{R}_2$ is consistent and $\text{At}(\mathcal{R}_1) \cap \text{At}(\mathcal{R}_2) = \emptyset$, then $\text{Inc}_\Pi^p(\mathcal{R}_1) = \text{Inc}_\Pi^p(\mathcal{R}_1 \cup \mathcal{R}_2)$.*

Proof. Let $\mathcal{P}_1$ be minimal violating for $\mathcal{R}_1$ and let $\mathcal{P}_2$ be a model of $\mathcal{R}_2$. Let $\Sigma_1 = \text{At}(\mathcal{R}_1)$ and $\Sigma_2 = \Sigma \setminus \Sigma_1$. For $i = 1, 2$, let $C_i = \{c \in \mathcal{L}_\Sigma \mid c = \bigwedge_{\alpha \in \Sigma_i} \alpha^{b_\alpha}, b_\alpha \in \{0, 1\}\}$ denote the set of complete conjunctions over Σ_i , where $\alpha^0 = \bar{\alpha}$ and $\alpha^1 = \alpha$ for each atom $\alpha \in \Sigma$. Note that $\Sigma = \Sigma_1 \uplus \Sigma_2$, hence we can identify each interpretation $\omega \in \Omega$ with the (unique) complete conjunction $c_1 c_2$ that is satisfied by ω , where $c_1 \in \Sigma_1, c_2 \in \Sigma_2$.

Now consider the function $\mathcal{P} : \Omega \rightarrow \mathbb{R}$ defined by $\mathcal{P}(c_1 c_2) = (\sum_{\bar{c}_2 \in \Sigma_2} \mathcal{P}_1(c_1 \bar{c}_2)) (\sum_{\bar{c}_1 \in \Sigma_1} \mathcal{P}_2(\bar{c}_1 c_2))$. As $\mathcal{P}_1$ and $\mathcal{P}_2$ are nonnegative, so is $\mathcal{P}$. Also, $\mathcal{P}(c_1 c_2) \leq 1$ by normalization of $\mathcal{P}_1$ and $\mathcal{P}_2$. Furthermore,

$$\begin{aligned} \mathcal{P}(\Omega) &= \sum_{c_1 \in \Sigma_1} \sum_{c_2 \in \Sigma_2} \mathcal{P}(c_1 c_2) \\ &= \underbrace{\sum_{c_1 \in \Sigma_1} \left(\sum_{\bar{c}_2 \in \Sigma_2} \mathcal{P}_1(c_1 \bar{c}_2) \right)}_{=1} \underbrace{\sum_{c_2 \in \Sigma_2} \left(\sum_{\bar{c}_1 \in \Sigma_1} \mathcal{P}_2(\bar{c}_1 c_2) \right)}_{=1} \\ &= 1, \end{aligned}$$

hence $\mathcal{P} : \Omega \rightarrow [0, 1]$ is normalized and therefore a probabilistic interpretation of $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$.

A formula F over Σ_1 is satisfied by $c_1 c_2$ if and only if F is satisfied by c_1 . Therefore,

$$\begin{aligned} \mathcal{P}(F) &= \sum_{c_1 \in \Sigma_1} \sum_{c_2 \in \Sigma_2} \mathcal{P}(c_1 c_2) 1_{\{F\}}(c_1) \\ &= \underbrace{\sum_{c_2 \in \Sigma_2} \left(\sum_{\bar{c}_1 \in \Sigma_1} \mathcal{P}_2(\bar{c}_1 c_2) \right)}_{=1} \underbrace{\sum_{c_1 \in \Sigma_1} \left(\sum_{\bar{c}_2 \in \Sigma_2} \mathcal{P}_1(c_1 \bar{c}_2) 1_{\{F\}}(c_1) \right)}_{=\mathcal{P}_1(F)} \\ &= \mathcal{P}_1(F). \end{aligned}$$

Hence, $\mathcal{P}$ coincides with $\mathcal{P}_1$ for formulas over Σ_1 and analogously it follows that $\mathcal{P}$ coincides with $\mathcal{P}_2$ for formulas over

Σ_2 . Therefore, if we denote by $\vec{\pi}, \vec{\pi}_1, \vec{\pi}_2$ the probability vectors corresponding to $\mathcal{P}, \mathcal{P}_1, \mathcal{P}_2$, it holds $\vec{a}_c^\top \vec{\pi} = \vec{a}_c^\top \vec{\pi}_1$ for all $c \in \mathcal{R}_1$ and $\vec{a}_c^\top \vec{\pi} = \vec{a}_c^\top \vec{\pi}_2 = 0$ for all $c \in \mathcal{R}_2$. Hence, if we number the conditionals in $\mathcal{R}_1$ by $1, \dots, m$ and the conditionals in $\mathcal{R}_2$ by $m+1, \dots, m+k$, we find

$$\begin{aligned} \min_{\vec{\pi} \in \mathbb{R}^n} \|A_{\mathcal{R}_1 \cup \mathcal{R}_2} \vec{\pi}\|_p &\leq \sqrt[p]{\sum_{i=1}^{m+k} |\vec{a}_i^\top \vec{\pi}|^p} \\ &= \sqrt[p]{\sum_{i=1}^m |\vec{a}_i^\top \vec{\pi}_1|^p + 0} = \text{Inc}_\Pi^p(\mathcal{R}). \end{aligned}$$

Hence, $\text{Inc}_\Pi^p(\mathcal{R}_1 \cup \mathcal{R}_2) \leq \text{Inc}_\Pi^p(\mathcal{R}_1)$. By monotonicity (Proposition 4.4), it also holds $\text{Inc}_\Pi^p(\mathcal{R}_1) \leq \text{Inc}_\Pi^p(\mathcal{R}_1 \cup \mathcal{R}_2)$ and therefore $\text{Inc}_\Pi^p(\mathcal{R}_1) = \text{Inc}_\Pi^p(\mathcal{R}_1 \cup \mathcal{R}_2)$. □

As a desirable property of inconsistency measures for probabilistic knowledge bases, Thimm (2009) considered *continuity*. The intuitive idea is that if we change the probabilities in a knowledge base slightly, the corresponding inconsistency value should change only slightly, too. Let $\mathcal{R}[\vec{\rho}] = \langle (\mathcal{G}_1 | \mathcal{F}_1)[\vec{\rho}_1], \dots, (\mathcal{G}_m | \mathcal{F}_m)[\vec{\rho}_m] \rangle$ denote an ordered knowledge base with conditional probabilities $\vec{\rho} \in \mathbb{R}^m$.

Proposition 4.8 (Continuity). *If $\lim_{k \rightarrow \infty} \vec{\rho}_k = \vec{\rho}$, then $\lim_{k \rightarrow \infty} \text{Inc}_\Pi^p(\mathcal{R}[\vec{\rho}_k]) = \text{Inc}_\Pi^p(\mathcal{R}[\vec{\rho}])$.*

Proof. Let $\mathcal{R} = \mathcal{R}[\vec{\rho}]$ be a knowledge base and let $(\vec{\rho}_k)$ be a sequence in $\mathbb{R}^m$ such that $\lim_{k \rightarrow \infty} \vec{\rho}_k = \vec{\rho}$. Let $A_{\mathcal{R}}$ be the matrix with row vectors $\vec{a}_i^\top$ corresponding to $\mathcal{R}$. Let A_k denote the matrix with row vectors $\vec{a}_{ki}^\top = (1_{\{\mathcal{G}_i | \mathcal{F}_i\}}(\omega_j) \cdot (1 - \vec{\rho}_{ki}) - 1_{\{\bar{\mathcal{G}}_i | \bar{\mathcal{F}}_i\}}(\omega_j) \cdot \vec{\rho}_{ki})_{1 \leq j \leq n}$ corresponding to $\mathcal{R}[\vec{\rho}_k]$. The linearity of the limit implies that $\lim_{k \rightarrow \infty} \vec{a}_{ki}^\top = \vec{a}_i^\top$ and that $\lim_{k \rightarrow \infty} \vec{a}_{ki}^\top \vec{\pi} = \vec{a}_i^\top \vec{\pi}$ for all $\vec{\pi} \in \mathbb{R}^n$. Hence, $\lim_{k \rightarrow \infty} A_k = A_{\mathcal{R}}$ and $\lim_{k \rightarrow \infty} A_k \vec{\pi} = A_{\mathcal{R}} \vec{\pi}$. Finally, by continuity of the p -norm, it follows that $\lim_{k \rightarrow \infty} \|A_k \vec{\pi}\|_p = \|A_{\mathcal{R}} \vec{\pi}\|_p$. As $\vec{\pi} \in \mathbb{R}^n$ was arbitrary, it follows $\lim_{k \rightarrow \infty} \text{Inc}_\Pi^p(\mathcal{R}[\vec{\rho}_k]) = \text{Inc}_\Pi^p(\mathcal{R}[\vec{\rho}])$. □

5 Computing Minimal Violation Measures

In this section, we explore how Inc_Π^p can be computed and how its computational performance relates to other measures. In general, computing Inc_Π^p corresponds to a convex optimization problem. Various approved algorithms exist to solve such problems, see (Boyd and Vandenberghe 2004) for an overview. For $p = 1$ and $p = \infty$, Inc_Π^p can be computed by linear programs, as I show in the following. The key idea of introducing auxiliary variables stems from unconstrained norm optimization problems (Boyd and Vandenberghe 2004).

For $p = 1$, we introduce a new vector $\vec{y} \in \mathbb{R}^m$ containing an auxiliary variable $\vec{y}_i$ for each conditional. Basically, the auxiliary variables $\vec{y}_i$ measure the error for the i -th conditional with respect to the 1-norm.

Proposition 5.1. $\text{Inc}_\Pi^1(\mathcal{R})$ can be computed by solving the following linear program:

$$\begin{aligned} \min_{(\vec{\pi}, \vec{y}) \in \mathbb{R}^{n+m}} \quad & \sum_{i=1}^m \vec{y}_i \quad (2) \\ \text{subject to} \quad & \sum_{i=1}^n \vec{\pi}_i = 1 \\ & -\vec{y} \leq A_{\mathcal{R}} \vec{\pi} \leq \vec{y} \\ & \vec{y} \geq 0 \\ & \vec{\pi} \geq 0. \end{aligned}$$

Proof. Along the same line of reasoning as for Proposition 3.3, it follows that (2) has a unique solution. It remains to show that optimization problem (2) yields the same solution as optimization problem (1) for $p = 1$ and arbitrary $\mathcal{R}$ over $(\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)$.

To begin with, let ϵ be an optimal solution of (1). Then there is a $\vec{\pi} \in \Pi_n$ such that $\epsilon = \|A_{\mathcal{R}} \vec{\pi}\|_1 = \sum_{i=1}^m |\vec{a}_i^T \vec{\pi}|$. Let $\vec{y} \in \mathbb{R}^m$ be defined by $\vec{y}_i = |\vec{a}_i^T \vec{\pi}|$. Then $(\vec{\pi}, \vec{y})$ is feasible for (2), hence its optimum is at most $\sum_{i=1}^m \vec{y}_i = \epsilon$.

Conversely, suppose that $\sum_{i=1}^m |\vec{y}_i|$ is optimal for (2) and that $(\vec{\pi}, \vec{y})$ is an optimal point. Then $\vec{\pi}$ is feasible for (1) and its optimum is at most $\|A_{\mathcal{R}} \vec{\pi}\|_1 = \sum_{i=1}^m |\vec{a}_i^T \vec{\pi}| \leq \sum_{i=1}^m |\vec{y}_i|$.

Hence both optima have to be equal. $\square$

For $p = \infty$, it suffices to introduce a single auxiliary variable $y \in \mathbb{R}$, because the maximum norm measures the maximum error with respect to all conditionals. In the following proposition, $y \vec{1}$ denotes multiplication of the vector $\vec{1} \in \mathbb{R}^m$ (that contains only ones) by the scalar y .

Proposition 5.2. $\text{Inc}_\Pi^\infty(\mathcal{R})$ can be computed by solving the following linear program:

$$\begin{aligned} \min_{(\vec{\pi}, y) \in \mathbb{R}^{n+1}} \quad & y \quad (3) \\ \text{subject to} \quad & \sum_{i=1}^n \vec{\pi}_i = 1 \\ & -y \vec{1} \leq A_{\mathcal{R}} \vec{\pi} \leq y \vec{1} \\ & y \geq 0 \\ & \vec{\pi} \geq 0. \end{aligned}$$

The proof is analogous to the last one and therefore left out.

Performance Evaluation

To test the computational performance of different measures, I computed inconsistency values for three test sets each of which contained nine knowledge bases and compared the runtime results. In each test set, from one knowledge base to the next, the number of conditionals and the size of the language was increased to capture the sensitivity of different measures with respect to the problem size. From

Test Set	$ \Sigma_i $	$ \Omega_i $	$ \mathcal{R}_i $
A	$2i$	4^i	$3i$
B	$2i$	4^i	3 for $i = 1$, $4i - 1$ for $i > 1$
C	$2i$	4^i	3 for $i = 1$, $5i - 2$ for $i > 1$

Table 1: Number of propositional variables $|\Sigma_i|$, worlds $|\Omega_i|$ and conditionals $|\mathcal{R}_i|$ in test sets A, B and C. For instance, $|\mathcal{R}_1^A| = 3$, $|\mathcal{R}_9^C| = 5 \cdot 9 - 2 = 43$, $|\Omega_9^C| = 4^9 = 2^{18}$.

one test set to the next, the number of conflicts between variables was increased to capture the sensitivity with respect to stronger correlations between the variables. The test sets were build up inductively as described in the following.

Test Set A: The first (inconsistent) knowledge base $\mathcal{R}_1^A$ contains three conditionals built up over two variables. Σ_i^A and $\mathcal{R}_i^A$ are obtained from Σ_{i-1}^A and $\mathcal{R}_{i-1}^A$ by adding a new copy of the variables in Σ_1^A and the conditionals in $\mathcal{R}_1^A$.

$$\begin{aligned} \Sigma_1^A &= \{A_1, B_1\}, \\ \mathcal{R}_1^A &= \{(A_1)[0.8], (B_1)[0.6], (B_1 | A_1)[0.8]\}, \\ \Sigma_i^A &= \Sigma_{i-1}^A \cup \{A_i, B_i\}, i > 1, \\ \mathcal{R}_i^A &= \mathcal{R}_{i-1}^A \cup \{(A_i)[0.8], (B_i)[0.6], (B_i | A_i)[0.8]\}, i > 1. \end{aligned}$$

Test Set B: For $i > 1$ the knowledge bases $\mathcal{R}_i^B$ differ from $\mathcal{R}_i^A$ by a new conflicting conditional that connects B_i and A_{i-1} . In this way all variables in test set B are correlated, whereas $\{A_i, B_i\}$ and $\{A_j, B_j\}$ are independent for $i \neq j$ in test set A.

$$\begin{aligned} \Sigma_i^B &= \Sigma_i^A, i \in \mathbb{N}, \\ \mathcal{R}_1^B &= \mathcal{R}_1^A, \\ \mathcal{R}_i^B &= \mathcal{R}_i^A \cup \{(B_k | A_{k-1})[0.85] \mid 2 \leq k \leq i\}, i > 1. \end{aligned}$$

Test Set C: Compared to test set B, test set C contains yet another conflicting conditional between consecutive sets.

$$\begin{aligned} \Sigma_i^C &= \Sigma_i^A, i \in \mathbb{N}, \\ \mathcal{R}_1^C &= \mathcal{R}_1^A, \\ \mathcal{R}_i^C &= \mathcal{R}_i^B \cup \{(A_i | B_{i-1})[0.1] \mid 2 \leq k \leq i\}, i > 1. \end{aligned}$$

Table 1 summarizes the calculation rules for the size of the knowledge bases and the corresponding languages.

Implementations: I used LPSolve¹, which is based on the Simplex algorithm, to compute Inc_Π^1 and Inc_Π^∞ . To compare the performance to other measures, I applied three prototypical implementations of the drastic measure (Hunter and Konieczny 2006; Thimm 2013), the MI measure (Hunter and Konieczny 2008; Thimm 2013) that counts the number of minimal inconsistent sets and the minimal change measure from (Thimm 2009).

¹<http://lpsolve.sourceforge.net>

The drastic measure was computed by Phase 1 of the Simplex algorithm, which basically searches for a feasible solution. State of the art solvers for the probabilistic satisfiability problem apply more sophisticated techniques to speed up the computation (Hansen and Perron 2008; Cozman and Ianni 2013), but these can be applied to compute Inc_{Π}^1 and $\text{Inc}_{\Pi}^{\infty}$ as well. Therefore, Phase 1 of the Simplex algorithm seems to be a reasonable baseline in this test series.

To find all minimal inconsistent sets, increasing subsets of the knowledge base were generated in lexicographic order. Each inconsistent subset was stored. The lexicographic order guaranteed minimality and that no supersets of minimal inconsistent sets were considered as candidates. To check consistency, again Phase 1 of the Simplex algorithm was applied. There already exist very sophisticated algorithms for enumerating minimal inconsistent sets for classical logics (Konieczny and Roussel 2013), but these rely on the fact that classical knowledge bases can be transformed into a single formula in conjunctive normal form. This fact, however, does not transfer to probabilistic logics.

The minimal change measure from (Thimm 2009) was computed by an implementation in the expert system shell MECore² (Finthammer et al. 2009). It is based on JCobyla³, a Java implementation of Cobyla that approximates the non-convex optimization problem by linear programs (Powell 1994). There already exists an implementation of the measure from (Thimm 2009) in the Tweety library⁴, but it relies on external libraries and is difficult to set up. Therefore, it was not used. Other methods might perform better than Cobyla, but to the best of my knowledge there are no algorithms for constrained non-convex problems that can handle more than a few thousand optimization variables and the used knowledge bases induced up to 2^{18} variables.

Figure 1 visualizes the runtime results. Note that the y-axis ranges to 10 seconds for test set A and B, but to 20 seconds for test set C. Execution was canceled if algorithms required more than 5 minutes to compute an inconsistency value.

Interpretation of Performance Results: The runtime for each algorithm in each test set in Figure 1 grows exponentially from one knowledge base to the next, because the number of optimization variables corresponds to the number of worlds $|\Omega|$ that is exponential in the size of $|\Sigma|$. Note that $|\Omega|$ quadruples from one knowledge base to the next one. Inc_{Π}^1 , $\text{Inc}_{\Pi}^{\infty}$ and the drastic measure scale much better than the minimal change measure and the minimal inconsistent set measure. In the next subsection, I summarize some empirical estimates from (Boyd and Vandenberghe 2004; Matousek and Gärtner 2007) for typical optimization algorithms that support these results.

Also note that the minimal inconsistent set measure is more sensitive with respect to the size of the knowledge

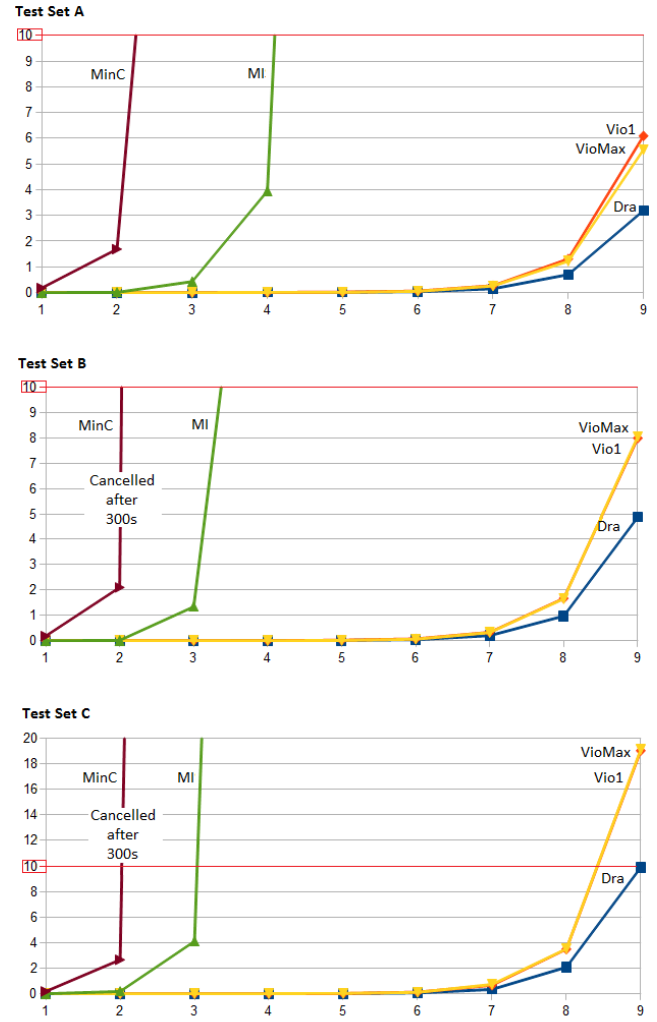


Figure 1: Runtime results in seconds (y-axis) for knowledge bases of increasing size (x-axis): Minimal change (MinC, red), minimal inconsistent sets (MI, green), Inc_{Π}^1 (Vio1, orange), $\text{Inc}_{\Pi}^{\infty}$ (VioMax, yellow), drastic (Dra, blue) measure.

base. For example, when comparing test set A and B, the runtime for linear programs increases by at most 50 percent for all knowledge bases, whereas the runtime for the minimal inconsistent set measure triples from $\mathcal{R}_3^A$ to $\mathcal{R}_3^B$ and quadruples from $\mathcal{R}_4^A$ to $\mathcal{R}_4^B$. This observation corresponds to the fact that the number of possible candidates for minimal inconsistent subsets increases exponentially with the size of the knowledge base.

On Computational Complexity

Finding reasonable estimates for the performance of inconsistency measures is difficult, because it depends heavily on the applied algorithm and the structure of the knowledge base at hand. For example, one can construct examples where the Simplex algorithm requires an exponential number of iterations, even though it often performs faster than

²The measure was implemented by Sebastian Lewalder in his ongoing bachelor thesis.

³<https://github.com/cureos/jcobyla>

⁴<http://tweety.sourceforge.net>

algorithms with a polynomial runtime guarantee in practice (Matousek and Gärtner 2007). In this section, I summarize some empirical estimates for typical optimization algorithms to provide some intuition for the practical difficulty of different measures.

Empirical estimates: Practical experiments indicate that linear programs can be solved by the Simplex algorithm in about $2\bar{m}$ to $3\bar{m}$ iterations (Matousek and Gärtner 2007), where $\bar{m}$ is the number of constraints after transforming the problem to a standard form consisting of equations and non-negativity constraints. Each iteration usually causes costs in the order of $|\Omega| |\mathcal{R}|$, so that the overall cost can be regarded as in the order of $|\Omega| |\mathcal{R}|^2$.

For convex programs, Boyd and Vandenberghe (2004) estimate 10 to 100 iterations with cost $\max\{|\Omega|^3, |\Omega|^2 |\mathcal{R}|\}$ per iteration when using interior-point methods. For the special case of $p = 2$, Inc_{Π}^2 corresponds to a quadratic programming problem that can be solved in the order of $|\Omega|^2 |\mathcal{R}|$ (Boyd and Vandenberghe 2004). Algorithms for non-convex programs can also be applied to convex programs, but usually are less efficient, because they rely on mechanisms to avoid (non-global) local minima. Therefore, one can expect that their computational cost is at least $|\Omega|^3$.

Practical Implications: As the number of worlds grows exponentially with the number of propositional variables, $|\Omega|$ usually dominates the size of the optimization problems. Therefore, one can expect to compute the drastic measure, Inc_{Π}^1 and $\text{Inc}_{\Pi}^{\infty}$ with linear complexity, Inc_{Π}^2 with quadratic complexity and Inc_{Π}^p for $p > 2$ with cubic complexity with respect to $|\Omega|$. For linear programs, several techniques can be applied to reduce the exponential influence of $|\Omega|$ (Jaumard, Hansen, and Poggi 1991; Hansen and Perron 2008; Cozman and Ianni 2013). The key idea is that only $\bar{m}$ columns of the $(\bar{m} \times \bar{n})$ constraint matrix have to be stored and in each iteration one column has to be switched. Selecting a good column from the $\bar{n} - \bar{m}$ possible candidates is the crucial problem.

Several measures rely on computing all minimal inconsistent sets (Hunter and Konieczny 2008; Thimm 2013). If $\#MI$ denotes the number of minimal inconsistent sets in a knowledge base, then we need to perform at least $\#MI$ satisfiability tests to verify their inconsistency. Based on the estimate $|\mathcal{R}|^2 |\Omega|$ for probabilistic satisfiability tests, the cost for computing minimal consistent sets is at least $\#MI |\mathcal{R}|^2 |\Omega|$.

As explained above, computing the non-convex problems corresponding to minimal change measures as introduced in (Thimm 2009) and (Picado-Muiño 2011) is at least as expensive as solving convex problems. Therefore, their computational cost can be expected to be greater than $|\Omega|^3$.

Table 2 summarizes the size of the optimization problems and the explained estimates for their practical computational cost. $2|\mathcal{R}|$ optimization variables for Inc_{Π}^1 and $\text{Inc}_{\Pi}^{\infty}$ correspond to slack variables that have to be generated to transform the problem into a normal form for the Simplex algorithm. For the same reason the number of constraints is increased.

Note, however, that the mere number of optimization variables $\bar{n}$ and constraints $\bar{m}$ can be misleading when compar-

Measure	$\bar{n}$	$\bar{m}$	Estimate
Drastic	$ \Omega + \mathcal{R} $	$ \mathcal{R} + 1$	$\sim \bar{n} \bar{m}^2$
$\text{Inc}_{\Pi}^{\infty}$	$ \Omega + 2 \mathcal{R} + 1$	$2 \mathcal{R} + 1$	$\sim \bar{n} \bar{m}^2$
Inc_{Π}^1	$ \Omega + 3 \mathcal{R} $	$2 \mathcal{R} + 1$	$\sim \bar{n} \bar{m}^2$
Inc_{Π}^2	$ \Omega $	1	$\sim \bar{n}^2 \bar{m}$
$\text{Inc}_{\Pi}^p, p > 2$	$ \Omega $	1	$\sim \bar{n}^3$
MI	$ \Omega + \mathcal{R} $	$ \mathcal{R} + 1$	$> \#MI \bar{n} \bar{m}^2$
MinChange	$ \Omega + \mathcal{R} $	$3 \mathcal{R} + 1$	$> \bar{n}^3$

Table 2: Number of optimization variables $\bar{n}$, number of constraints $\bar{m}$ (without non-negativity constraints) and rough performance estimates for different measures. $\#MI$ denotes the number of minimal inconsistent sets. Performance estimates are based on empirical values from (Boyd and Vandenberghe 2004; Matousek and Gärtner 2007) for typical optimization algorithms.

ing two algorithms. For example, as Table 2 shows, the overall number of variables $\bar{n}$ is smaller for $\text{Inc}_{\Pi}^{\infty}$ than for Inc_{Π}^1 . But whereas the single auxiliary variable for $\text{Inc}_{\Pi}^{\infty}$ appears in $2|\mathcal{R}|$ constraints, each of the $|\mathcal{R}|$ auxiliary variables for Inc_{Π}^1 appear in only 2 constraints (c.f. Proposition 5.1 and 5.2). Therefore, the variables are looser correlated for Inc_{Π}^1 and, indeed, $\text{Inc}_{\Pi}^{\infty}$ is slightly outperformed by Inc_{Π}^1 in Figure 1 for test sets B and C.

6 Conclusion and Future Work

As shown in Section 4, minimal violation measures satisfy many desirable properties that have been considered in the literature (Knight 2003; Hunter and Konieczny 2006; Thimm 2013). In contrast to the naive drastic measure or measures based on minimal inconsistent sets (Hunter and Konieczny 2008; Thimm 2013), minimal violation measures take the numerical structure of probabilistic knowledge bases into account and in this way provide a more fine-grained insight into inconsistencies, as demonstrated in Examples 3.4 and 3.5. When applied as a guide to repair knowledge bases, they can never yield better theoretical guarantees than measures based on minimal changes, because these yield the best repair by definition (Thimm 2009; Picado-Muiño 2011). However, the lack of efficient algorithms for constrained non-convex problems prevents us from computing minimal change measures for larger knowledge bases. As explained in Section 5, minimal violation measures possess better computational properties. In particular, Inc_{Π}^1 and $\text{Inc}_{\Pi}^{\infty}$ stand out among conventional measures for probabilistic logics, because they provide similar good performance guarantees like the naive drastic measure, while measuring the degree of inconsistency continuously.

One long term goal of this work is resolving inconsistencies in probabilistic knowledge bases with methods that are better guided than purely heuristic approaches and that provide better computational properties than minimal change approaches. A first algorithm based on minimal violation measures might iterate over the conditionals, check whether an adaptation yields an improvement in the inconsistency

value and select the best adaptation as illustrated in Example 3.5. To speed up repeated computations of Inc_{Π}^1 and $\text{Inc}_{\Pi}^{\infty}$, column generation techniques can be applied (Hansen and Perron 2008). In future work such an algorithm shall be implemented and evaluated in the expert system shell MECore (Finthammer et al. 2009).

It might also be interesting to ask whether the minimal violating interpretations can be used to infer reasonable information from contradictory knowledge bases. The minimal violating interpretations yield a compact and convex set just like the models of consistent knowledge bases. Therefore, computing probabilities from minimal violating interpretations of inconsistent knowledge bases can be performed like computing probabilities from the models of consistent knowledge bases. However, the question remains whether the computed probabilities yield reasonable results and whether they can be justified by some rationales.

For $p = 1$, $p = 2$ and $p = \infty$, Java-implementations of Inc_{Π}^p are made available in the Tweety library⁵ (they can be found in the package `net.sf.tweety.logics.pcl.analysis`. Note that LPSolve⁶ has to be installed and configured properly to use Inc_{Π}^1 and $\text{Inc}_{\Pi}^{\infty}$).

References

- Boyd, S., and Vandenberghe, L. 2004. *Convex Optimization*. New York, NY, USA: Cambridge University Press.
- Cozman, F., and Ianni, L. 2013. Probabilistic satisfiability and coherence checking through integer programming. In van der Gaag, L. C., ed., *Proc. Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2013*, volume 7958 of *Lecture Notes in Computer Science*, 145–156. Springer Berlin Heidelberg.
- Doder, D.; Raskovic, M.; Markovic, Z.; and Ognjanovic, Z. 2010. Measures of inconsistency and defaults. *International Journal of Approximate Reasoning* 51(7):832–845.
- Finthammer, M.; Beierle, C.; Berger, B.; and Kern-Isberner, G. 2009. Probabilistic reasoning at optimum entropy with the MECore system. In Lane, H., and Guesgen, H., eds., *Proc. International FLAIRS Conference, FLAIRS 2009*, 535–540. AAAI Press, Menlo Park, California.
- Finthammer, M.; Kern-Isberner, G.; and Ritterskamp, M. 2007. Resolving inconsistencies in probabilistic knowledge bases. In Hertzberg, J.; Beetz, M.; and Englert, R., eds., *Proc. German Conference on Artificial Intelligence, KI 2007*, volume 4667 of *Lecture Notes in Computer Science*, 114–128. Springer.
- Georgakopoulos, G. F.; Kavvadias, D. J.; and Papadimitriou, C. H. 1988. Probabilistic satisfiability. *Journal of Complexity* 4(1):1–11.
- Grant, J., and Hunter, A. 2008. Analysing inconsistent first-order knowledgebases. *Artificial Intelligence* 172(8-9):1064–1093.
- Hansen, P., and Perron, S. 2008. Merging the local and global approaches to probabilistic satisfiability. *International Journal of Approximate Reasoning* 47(2):125 – 140.
- Hunter, A., and Konieczny, S. 2006. Shapley inconsistency values. In Doherty, P.; Mylopoulos, J.; and Welty, C. A., eds., *Proc. Principles of Knowledge Representation and Reasoning, KR 2006*, 249–259. AAAI Press.
- Hunter, A., and Konieczny, S. 2008. Measuring inconsistency through minimal inconsistent sets. In Brewka, G., and Lang, J., eds., *Proc. Principles of Knowledge Representation and Reasoning, KR 2008*, 358–366. AAAI Press.
- Jaumard, B.; Hansen, P.; and Poggi, M. 1991. Column generation methods for probabilistic logic. *ORSA - Journal on Computing* 3(2):135–148.
- Kern-Isberner, G. 2001. *Conditionals in Nonmonotonic Reasoning and Belief Revision*. Number 2087 in *Lecture Notes in Computer Science*. Berlin: Springer-Verlag.
- Knight, K. 2002. Measuring inconsistency. *Journal of Philosophical Logic* 31(1):77–98.
- Knight, K. M. 2003. Two information measures for inconsistent sets. *Journal of Logic, Language and Information* 12(2):227–248.
- Konieczny, S., and Roussel, S. 2013. A reasoning platform based on the MI shapley inconsistency value. In van der Gaag, L. C., ed., *Proc. Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2013*, volume 7958 of *Lecture Notes in Computer Science*, 315–327. Springer.
- Lukasiewicz, T. 1999. Probabilistic deduction with conditional constraints over basic events. *Journal of Artificial Intelligence Research* 10:380–391.
- Matousek, J., and Gärtner, B. 2007. *Understanding and Using Linear Programming*. Universitext (1979). Springer.
- Nilsson, N. J. 1986. Probabilistic logic. *Artificial Intelligence* 28:71–88.
- Paris, J. B., and Vencovská, A. 1989. On the applicability of maximum entropy to inexact reasoning. *International Journal of Approximate Reasoning* 3(1):1–34.
- Picado-Muiño, D. 2011. Measuring and repairing inconsistency in probabilistic knowledge bases. *International Journal of Approximate Reasoning* 52(6):828–840.
- Powell, M. 1994. A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in Optimization and Numerical Analysis*, volume 275 of *Mathematics and Its Applications*. Springer Netherlands. 51–67.
- Rödder, W., and Xu, L. 1999. Entropy-driven inference and inconsistency. In *Proc. Artificial Intelligence and Statistics*, volume 99, 272–277.
- Thimm, M. 2009. Measuring inconsistency in probabilistic knowledge bases. In Bilmes, J., and Ng, A. Y., eds., *Proc. Uncertainty in Artificial Intelligence, UAI 2009*, 530–537. AUAI Press.
- Thimm, M. 2013. Inconsistency measures for probabilistic logics. *Artificial Intelligence* 197:1–24.

⁵<http://tweety.sourceforge.net>

⁶<http://lpsolve.sourceforge.net>