Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference

A Novel Methodology for Processing Probabilistic Knowledge BasesUnder Maximum Entropy

Gabriele Kern-Isberner and **Marco Wilhelm**
TU Dortmund University
Dortmund, Germany

Christoph Beierle
University of Hagen
Hagen, Germany

Abstract

Probabilistic reasoning under the so-called principle of maximum entropy is a viable and convenient alternative to Bayesian networks, relieving the user from providing complete (local) probabilistic information and observing rigorous conditional independence assumptions. In this paper, we present a novel approach to performing computational MaxEnt reasoning that makes use of symbolic computations instead of graph-based techniques. Given a probabilistic knowledge base, we encode the MaxEnt optimization problem into a system of polynomial equations, and then apply Gröbner basis theory to find MaxEnt inferences as solutions to the polynomials. We illustrate our approach with an example of a knowledge base that represents findings on fraud detection in enterprises.

1 Introduction

Probability theory provides one of the richest and most popular frameworks for uncertain reasoning with efficient graph-based propagation techniques like Bayesian networks (cf., e.g., (Cowell et al. 1999; Pearl 1988)). However, probabilistic reasoning is problematic if the available information is incomplete, e.g., the Bayesian network approach does not work in such cases. Moreover, the rigorous conditional independence assumptions that are indispensable for Bayesian networks may be deemed inappropriate in general. The *principle of maximum entropy* (in short, *MaxEnt principle*) offers an alternative for probabilistic reasoning that overcomes these weaknesses of Bayesian networks – it relies on specified conditional dependencies, and as an inductive reasoning method, it completes incomplete information in a most cautious way (Shore and Johnson 1980; Jaynes 1983), yielding unique probability distributions from probabilistic knowledge bases, and matches the ideas of probabilistic commonsense reasoning perfectly (Paris 1999). For computing MaxEnt distributions, efficient tools can be used (Rödder and Meyer 1996). Nevertheless, in spite of its proved excellent general properties (Paris 1994; Kern-Isberner 2001), MaxEnt reasoning is still perceived as a black box methodology that returns probabilities according to some abstract optimization principle.

In this paper, we present a novel methodology for performing probabilistic reasoning at maximum entropy that is based on the strong conditional-logical structures that underly the MaxEnt principle and makes use of symbolic computations to process information in a generic way. We show how MaxEnt distributions can be obtained by solving systems of polynomial equations in which probabilities occur as symbolic parameters. This provides new insights into MaxEnt reasoning by abstracting from numerical peculiarities and representing the dependencies between the probabilistic rules in the knowledge base in an algebraic way. The methodology of Gröbner bases (Buchberger 2006; Cox, Little, and O'Shea 2007) from computer algebra can then be applied to perform computational probabilistic reasoning according to the MaxEnt principle, and to provide answers for queries to the knowledge base, yielding not only the inferred probabilities but revealing also the conditional-logical grounds on which the inference is based with respect to the given knowledge base. We present the basic ideas of our approach (for a more detailed mathematical elaboration see (Kern-Isberner, Wilhelm, and Beierle 2014)) and illustrate them with an example from auditing in which evidence for fraud (so-called *red flags*) can be combined to yield an overall estimation how probable fraud is in the enterprise under consideration, building on previous work (Finthammer, Kern-Isberner, and Ritterskamp 2007).

The organization of this paper is as follows: Section 2 gives a short recall of probabilistic knowledge representation and the MaxEnt principle. Section 3 provides an insight into computer algebra with Gröbner bases. Our approach of combining Gröbner basis methods and MaxEnt reasoning is presented in Section 4. Section 5 shows how to answer MaxEnt queries symbolically. In Section 6, an example in the field of auditing illustrates the presented methodology. Section 7 concludes the paper with a short summary and an outlook.

2 Basics of Knowledge Representation

Consider a probabilistic conditional language $(\mathcal{L}|\mathcal{L})^{prob} = \{(B|A)[x] \,|\, (B|A) \in (\mathcal{L}|\mathcal{L}), x \in [0,1]\}$ with Roman uppercase letters denoting atoms or formulas in a propositional language $\mathcal{L}$ over a finite alphabet. The language $\mathcal{L}$ is equipped with the common logical connectives $\wedge$ (and), $\vee$ (or) and $\neg$ $(negation)$. To shorten mathematical formulas, we write AB instead of $A \wedge B$ and $\overline{A}$ instead of $\neg A$. An element $(B|A)[x] \in (\mathcal{L}|\mathcal{L})^{prob}$, called *(probabilistic) conditional*, may be understood as the phrase "If A, then B with probability x". Formally, we have to introduce the concept of a probability distribution $\mathcal{P}$ on $\mathcal{L}$. Therefore, let Ω be the set of all possible worlds ω; here, Ω is simply a complete set of interpretations of $\mathcal{L}$. If a world ω satisfies a formula (or atom) A, we write $\omega \models A$ and call ω a *model* of A. Usually, we identify each possible world ω with the minterm (or complete conjunction) that has exactly ω as a model. Then, every $A \in \mathcal{L}$ can be assigned a probability via $\mathcal{P}(A) = \sum_{\omega \models A} \mathcal{P}(\omega)$. Conditionals are interpreted by distributions via conditional probabilities. If $\mathcal{P}$ is a probability distribution on Ω resp. $\mathcal{L}$, satisfaction of a conditional by $\mathcal{P}$ is defined by $\mathcal{P} \models (B|A)[x]$ iff $\mathcal{P}(A) > 0$ and $x = \mathcal{P}(B|A) = \frac{\mathcal{P}(AB)}{A}$. A probability distribution $\mathcal{P}$ satisfies a set of conditionals $\mathcal{C} \subseteq (\mathcal{L}|\mathcal{L})^{prob}$ iff $\mathcal{P}$ satisfies every conditional in $\mathcal{C}$. The set $\mathcal{C}$ is called *consistent* iff there exists a distribution satisfying it. A finite set of conditionals $\mathcal{KB} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\} \subseteq (\mathcal{L}|\mathcal{L})^{prob}$ is called a *knowledge base*, and there may exist several (or none) distributions satisfying it since usually, $\mathcal{KB}$ represents incomplete knowledge. In order to use inductively the information in $\mathcal{KB}$ it is very helpful to choose a "best" model of $\mathcal{KB}$. The *principle of maximum entropy* (cf. (Kern-Isberner 1998) and (Paris 1994)) provides a well-known solution to this problem by fulfilling the paradigm of *informational economy*, i.e., of least amount of assumed information (cf. (Gärdenfors 1988)). Therefore, one maximizes the entropy $H(\mathcal{Q}) = -\sum_{\omega \in \Omega} \mathcal{Q}(\omega) \log \mathcal{Q}(\omega)$ of a distribution $\mathcal{Q}$ with $\mathcal{Q}$ being a model of $\mathcal{KB}$. It can be shown that for every consistent knowledge base $\mathcal{KB}$ such a distribution $\mathcal{ME}(\mathcal{KB})$ with maximal entropy exists, and, in particular, $\mathcal{ME}(\mathcal{KB})$ is unique (cf. (Paris 1994)). It immediately follows that $\mathcal{KB}$ is consistent iff $\mathcal{ME}(\mathcal{KB})$ exists. Taking the conventions $\infty^0 = 1$, $\infty^{-1} = 0$ and $0^0 = 1$ into account, the distribution $\mathcal{ME}(\mathcal{KB})$ is given by

$$\mathcal{ME}(\mathcal{KB})(\omega) = \alpha_0 \prod_{\substack{1 \le i \le n \\ \omega \models A_i B_i}} \alpha_i^{1-x_i} \prod_{\substack{1 \le i \le n \\ \omega \models A_i \overline{B_i}}} \alpha_i^{-x_i} \tag{1}$$

with a normalizing constant α_0 and *effects* $\alpha_i > 0$ iff $x_i \in (0, 1)$, $\alpha_i = \infty$ iff $x_i = 1$, and $\alpha_i = 0$ iff $x_i = 0$. The effects α_i are associated with the corresponding conditionals and solve the following system of non-linear equations

$$\begin{aligned}(1 - x_i)\, \alpha_i^{1-x_i} \sum_{\omega \models A_i B_i} \prod_{\substack{j \ne i \\ \omega \models A_j B_j}} \alpha_j^{1-x_j} \prod_{\substack{j \ne i \\ \omega \models A_j \overline{B_j}}} \alpha_j^{-x_j} \\ = x_i\, \alpha_i^{-x_i} \sum_{\omega \models A_i \overline{B_i}} \prod_{\substack{j \ne i \\ \omega \models A_j B_j}} \alpha_j^{1-x_j} \prod_{\substack{j \ne i \\ \omega \models A_j \overline{B_j}}} \alpha_j^{-x_j}\end{aligned} \tag{2}$$

for $1 \le i \le n$ (cf. (Kern-Isberner 2001)); note that the α_i follow the three-valued logics of conditionals, being ineffective on $\overline{A_i}$. If $\mathcal{ME}(\mathcal{KB})$ exists, we can compute the MaxEnt probability of any further conditional $(B|A)$ from $\mathcal{ME}(\mathcal{KB})$. This yields a (non-monotonic) *MaxEnt inference relation* $\mid\!\sim_{\mathcal{ME}}$ with $\mathcal{KB} \mid\!\sim_{\mathcal{ME}} (B|A)[x]$ iff $\mathcal{ME}(\mathcal{KB}) \models (B|A)[x]$.

3 Basics of Gröbner Basis Theory

Gröbner bases are specific generating sets of ideals in polynomial rings that allow to condense information given by algebraic specifications of problems; a recommendable reference for the material presented in this section is (Cox, Little, and O'Shea 2007). Let $\mathbb{Q}[\mathcal{Y}]$ be the polynomial ring in variables $\mathcal{Y} = \{y_0, y_1, \ldots, y_s\}$ over the field of rational numbers $\mathbb{Q}$. Polynomials $f \in \mathbb{Q}[\mathcal{Y}]$ may be understood as finite linear combinations of terms over $\mathbb{Q}$ where a *term* is an element of the set $\mathcal{T} = \{y_0^{e_0}\, y_1^{e_1} \cdots y_s^{e_s} \mid e_0, e_1, \ldots, e_s \in \mathbb{N}_0\}$. The set of terms occurring in a polynomial $f \in \mathbb{Q}[\mathcal{Y}]$ with non-vanishing coefficients is called *support* of f, written $supp(f)$. An element $(\zeta_0, \zeta_1, \ldots, \zeta_s) \in \mathbb{C}^{s+1}$ is called a *root* of f iff $f(\zeta_0, \zeta_1, \ldots, \zeta_s) = 0$. Terms $t \in \mathcal{T}$ can be embedded into $\mathbb{Q}[\mathcal{Y}]$ as monomials with coefficient 1. Differently from the univariate case, terms in several variables can be ordered (reasonably) in many ways.

Definition 1 (Term Ordering, $lc_\preceq$, $lm_\preceq$). *Let $\preceq$ be a total ordering with related strict ordering $\prec$ on $\mathcal{T}$. $\preceq$ is a* term ordering *iff for all $t \in \mathcal{T}$ we have $1 = y_0^0\, y_1^0 \cdots y_s^0 \preceq t$, and for all $t, t_1, t_2 \in \mathcal{T}$, $t_1 \preceq t_2$ implies $t\, t_1 \preceq t\, t_2$.*

Let $\preceq$ be a term ordering on $\mathcal{T}$ and $f = \sum_{i=1}^{m} c_i\, t_i \in \mathbb{Q}[\mathcal{Y}]$ with $c_i \in \mathbb{Q} \setminus \{0\}$, $t_i \in \mathcal{T}$ for $1 \le i \le m$ and $t_1 \prec \ldots \prec t_m$. The leading coefficient *of f is $lc_\preceq(f) = c_m$, and the* leading monomial *of f is $lm_\preceq(f) = c_m\, t_m$.*

An important class of term orderings are the so called *elimination term orderings*. With elimination term orderings, it is possible to expose the part of an ideal that depends on certain variables, only. A term ordering $\preceq$ on $\mathcal{T}$ is an elimination term ordering for the variables $y_1, \ldots, y_s$ iff for all $f \in \mathbb{Q}[\mathcal{Y}]$, $lm_\preceq(f) \in \mathbb{Q}[y_0]$ implies $f \in \mathbb{Q}[y_0]$. An example of such a term ordering is the *lexicographical term ordering* $\preceq_{\text{lex}}$ on $\mathcal{T}$ that is recursively defined by $y_0^{e_0}\, y_1^{e_1} \cdots y_s^{e_s} \prec_{\text{lex}} y_0^{f_0}\, y_1^{f_1} \cdots y_s^{f_s}$ iff $e_s < f_s$ or $e_s = f_s$ and $y_0^{e_0}\, y_1^{e_1} \cdots y_{s-1}^{e_{s-1}} \prec_{\text{lex}} y_0^{f_0}\, y_1^{f_1} \cdots y_{s-1}^{f_{s-1}}$ presupposing that $y_0 \prec_{\text{lex}} y_1 \prec_{\text{lex}} \ldots \prec_{\text{lex}} y_s$. In particular, such an elimination term ordering always exists.

Definition 2 (Ideal). *A subset $\mathcal{I} \subseteq \mathbb{Q}[\mathcal{Y}]$ is called a* (polynomial) ideal *iff $0 \in \mathcal{I}$ and for all $f, g \in \mathcal{I}$, $h \in \mathbb{Q}[\mathcal{Y}]$ also $f + g \in \mathcal{I}$ as well as $h\, f \in \mathcal{I}$.*

Let $\mathcal{I} \subseteq \mathbb{Q}[\mathcal{Y}]$ be an ideal, and let $\mathcal{F} \subseteq \mathcal{I}$ so that for all $f \in \mathcal{I}$ there are $f_1, \ldots, f_m \in \mathcal{F}$ and $h_1, \ldots, h_m \in \mathbb{Q}[\mathcal{Y}]$ such that $f = \sum_{i=1}^{m} h_i\, f_i$ holds. Then $\mathcal{F}$ is called a *generating set* of $\mathcal{I}$, written $\mathcal{I} = \langle \mathcal{F} \rangle$. Obviously, the ideal $\mathcal{I}$ only consists of polynomials that vanish in the common roots of the polynomials in $\mathcal{F}$. Therefore, we can speak of the common roots of $\mathcal{I}$, which are exactly the same as the common roots of $\mathcal{F}$.

In order to understand the fundamental importance of Gröbner bases, imagine that the problem under consideration can be described by a set $\mathcal{F}$ of polynomials, and the solutions of the problem correspond to the common roots of $\mathcal{F}$. The ideal generated by $\mathcal{F}$ provides an algebraic context to condense the problem description without changing (essentially) the solutions of the problem.

Definition 3 (Gröbner Basis). *Let $\mathcal{I} \subseteq \mathbb{Q}[\mathcal{Y}]$ be an ideal with $\mathcal{I} \neq \langle \{0\} \rangle$ and let $\preceq$ be a term ordering on $\mathcal{T}$.*

A subset $\mathcal{B}_\preceq = \{b_1, \ldots, b_m\} \subseteq \mathcal{I}$ *is called a* Gröbner basis *for* $\mathcal{I}$ *with respect to* $\preceq$ *iff* $\langle\{lm_\preceq(b) \,|\, b \in \mathcal{B}_\preceq\}\rangle = \langle\{lm_\preceq(f) \,|\, f \in \mathcal{I}\}\rangle$. *In particular,* $\mathcal{I} = \langle\mathcal{B}_\preceq\rangle$ *holds in this case, i.e.,* $\mathcal{B}_\preceq$ *is a generating set of* $\mathcal{I}$. $\mathcal{B}_\preceq$ *is called a* minimal *Gröbner basis for* $\mathcal{I}$ *with respect to* $\preceq$ *iff in addition* $lc(b_i) = 1$ *and* $t \notin \langle\{lm(\mathcal{B}_\preceq \setminus \{b_i\})\}\rangle$ *hold for all* $t \in supp(b_i)$ *and* $1 \leq i \leq m$.

As a consequence of the *Hilbert's Basis Theorem*, every ideal $\mathcal{I} \subseteq \mathbb{Q}[\mathcal{Y}]$ with $\mathcal{I} \neq \langle\{0\}\rangle$ has a *unique* minimal Gröbner basis with respect to a given term ordering $\preceq$, written $\mathcal{GB}_\preceq(\mathcal{I})$ (cf. (Cox, Little, and O'Shea 2007)). The standard method to calculate Gröbner bases is *Buchberger's algorithm* that is implemented in all current computer algebra systems such as Maple or Mathematica.

Given an ideal $\mathcal{I} \subseteq \mathbb{Q}[\mathcal{Y}]$, it is possible to focus on just one variable, say y_0. The intersection $\mathcal{I} \cap \mathbb{Q}[y_0]$ is still an ideal, called *elimination ideal* of $\mathcal{I}$ for $y_1, \ldots, y_s$. In order to determine $\mathcal{I} \cap \mathbb{Q}[y_0]$, one derives a minimal Gröbner basis for $\mathcal{I}$ with respect to an elimination term ordering for $y_1, \ldots, y_s$. By deleting all polynomials with terms containing at least one of the variables $y_1, \ldots, y_s$, one obtains a Gröbner basis for $\mathcal{I} \cap \mathbb{Q}[y_0]$. Note that elimination ideals are usually defined more generally which is not necessary in our case. To gain a first insight, again (Cox, Little, and O'Shea 2007) is recommended.

4 Polynomial Representation of Probabilistic Knowledge Bases

Let $\mathcal{KB} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$ be a probabilistic conditional knowledge base. For our further considerations in this paper we assume the probabilities x_i to be rational numbers and non-trivial, i.e., $x_i \in (0,1)$ for $1 \leq i \leq n$. Thus, for each probability x_i, there are unique natural numbers $p_i, q_i \in \mathbb{N}$ such that $\frac{p_i}{q_i} = x_i$ and p_i, q_i are relatively prime. Indeed, conditional probability constraints in practical applications are usually rational, and cases where $x_i = 0$ resp. $x_i = 1$ for some $1 \leq i \leq n$ occur can be treated similarly, or even more simply (cf. (Kern-Isberner 2001)). For applying Gröbner bases methods to knowledge representation, it is necessary to transform the system of equations (2) into a *polynomial* equation system. Since the probabilities are assumed to be rational, we may apply the substitution

$$y_i^{q_i} := \alpha_i \tag{3}$$

for $1 \leq i \leq n$ to (2). Multiplying both sides with $q_i\, y_i^{p_i} \prod_{j \neq i} y_j^{p_j}$ and rearranging terms lead to $f_i = 0$ with

$$\begin{aligned} f_i := (q_i - p_i)\, y_i^{q_i} \sum_{\omega \models A_i B_i} \prod_{\substack{j \neq i \\ \omega \models A_j B_j}} y_j^{q_j} \prod_{\substack{j \neq i \\ \omega \models \overline{A_j}}} y_j^{p_j} \\ - p_i \sum_{\omega \models A_i \overline{B_i}} \prod_{\substack{j \neq i \\ \omega \models A_j B_j}} y_j^{q_j} \prod_{\substack{j \neq i \\ \omega \models \overline{A_j}}} y_j^{p_j} \end{aligned} \tag{4}$$

for $1 \leq i \leq n$. Then, $\mathcal{F} := \{f_1, \ldots, f_n\}$ is a set of polynomials in the variables $y_1, \ldots, y_n$ which represent the original conditionals in the knowledge base according to (2) and (3).

As vanishing the polynomials in (4) describes a necessary condition for the effects α_i of the conditionals in $\mathcal{KB}$ (more precisely for α_i^{1/q_i}), we are interested in the (real and positive) common roots of $\mathcal{F}$. Note that $\alpha_i = 0$ and therefore $y_i = 0$ iff $x_i = 0$ for $1 \leq i \leq n$. As we concentrate on knowledge bases with non-trivial probabilities $x_i \in (0,1)$, the accomplished transformation of (2) does not mean any loss of information, and we may ignore trivial roots of (4), i.e., roots with at least one entry that is zero. Therefore, we cancel out variables if possible, i.e., we repeatedly divide f_i by y_j for $1 \leq i, j \leq n$ until the result is still a polynomial. Furthermore, we may cancel polynomial combinations of variables if all of the coefficients are positive. Since this applies to both polynomial combinations

$$f_i^+ := (q_i - p_i)\, y_i^{q_i} \sum_{\omega \models A_i B_i} \prod_{\substack{j \neq i \\ \omega \models A_j B_j}} y_j^{q_j} \prod_{\substack{j \neq i \\ \omega \models \overline{A_j}}} y_j^{p_j},$$

$$f_i^- := p_i \sum_{\omega \models A_i \overline{B_i}} \prod_{\substack{j \neq i \\ \omega \models A_j B_j}} y_j^{q_j} \prod_{\substack{j \neq i \\ \omega \models \overline{A_j}}} y_j^{p_j},$$

we divide the polynomial f_i by the *greatest common divisor* $\gcd(f_i^+, f_i^-)$. Note that $\gcd(f_i^+, f_i^-)$ is positive for any assignment of $y_1, \ldots, y_n$ which leads to the effects of the conditionals in $\mathcal{KB}$. The result is still a polynomial. Altogether, we observe the set of polynomials $\mathcal{F}^* := \{f_1^*, \ldots, f_n^*\}$ with

$$f_i^* := \frac{f_i^+ - f_i^-}{\gcd(f_i^+, f_i^-)} \tag{5}$$

for $1 \leq i \leq n$. In (Cox, Little, and O'Shea 2007), it can be found how the greatest common divisor of multivariate polynomials can be derived using Gröbner bases methods.

As a first result of applying Gröbner bases methods to reasoning under the MaxEnt principle, we formulate a necessary condition for the consistency of a knowledge base. Note that Theorem 1 is a refinement of Theorem 2 in (Kern-Isberner, Wilhelm, and Beierle 2014).

Theorem 1. *Let* $\mathcal{KB} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$ *be a consistent knowledge base with non-trivial probabilities and let* $\preceq$ *be a term odering on* $\mathcal{T}$. *Then* $\mathcal{GB}_\preceq(\langle\mathcal{F}^*\rangle) \neq \{1\}$.

5 MaxEnt Reasoning for Answering Queries

For our further investigations, it is essential to know what inferences can be drawn from a consistent knowledge base $\mathcal{KB} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$ under the MaxEnt methodology. So, let $(B|A)$ be an (additional) arbitrary conditional. Then, $\mathcal{KB} \mid\!\sim_{\mathcal{ME}} (B|A)[x]$ is satisfied iff

$$x = \frac{\mathcal{ME}(\mathcal{KB})(AB)}{\mathcal{ME}(\mathcal{KB})(A)}. \tag{6}$$

Hence, if $\mathcal{ME}(\mathcal{KB})$ is known, it is possible to derive x from (6). To apply Gröbner bases methods, it is necessary to formulate a polynomial counterpart for (6). Therefore, we associate the variable y_0 with the unknown probability x. Making use of (1) and the substitutions $\frac{p_i}{q_i} = x_i$ as well as (3) for $1 \leq i \leq n$, (6) leads to the new equation $\hat{f} = 0$ with the polynomial

$$\hat{f} := y_0 \sum_{\omega \models A} \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i B_i}} y_i^{q_i} \prod_{\substack{1 \leq i \leq n \\ \omega \models \overline{A_i}}} y_i^{p_i} - \sum_{\omega \models AB} \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i B_i}} y_i^{q_i} \prod_{\substack{1 \leq i \leq n \\ \omega \models \overline{A_i}}} y_i^{p_i} \tag{7}$$

in the variables $y_0, y_1, \ldots, y_n$. Since all coefficients of the expressions

$$\hat{f}^+ := y_0 \sum_{\omega \models A} \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i B_i}} y_i^{q_i} \prod_{\substack{1 \leq i \leq n \\ \omega \models \overline{A_i}}} y_i^{p_i}$$

$$\text{and} \quad \hat{f}^- := \sum_{\omega \models AB} \prod_{\substack{1 \leq i \leq n \\ \omega \models A_i B_i}} y_i^{q_i} \prod_{\substack{1 \leq i \leq n \\ \omega \models \overline{A_i}}} y_i^{p_i}$$

are 1, and thus, they are positive, we may divide $\hat{f}$ through $\gcd(\hat{f}^+, \hat{f}^-)$ similar to (5). The resulting polynomial is

$$\hat{f}^* := \frac{\hat{f}^+ - \hat{f}^-}{\gcd(\hat{f}^+, \hat{f}^-)}. \tag{8}$$

The next theorem derives a necessary condition for the MaxEnt probability x.

Theorem 2 (MaxEnt Inference). *Given a consistent knowledge base $\mathcal{KB} = \{(B_1|A_1)[x_1], \ldots, (B_n|A_n)[x_n]\}$ with non-trivial probabilities and a further conditional $(B|A)$ with $\mathcal{KB} \mid\!\sim_{\mathcal{ME}} (B|A)[x]$, the MaxEnt probability x is a common root of $\langle \mathcal{F}^* \cup \{\hat{f}^*\} \rangle \cap \mathbb{Q}[y_0]$.*

Proof. As $\mathcal{KB}$ is consistent with non-trivial probabilities, $\mathcal{ME}(\mathcal{KB})$ exists and there is a common root $(\zeta_1, \ldots, \zeta_n) \in (0, \infty)^n$ of $\mathcal{F}$ and also of $\mathcal{F}^*$. The MaxEnt probability x is given then by (6) and $(x, \zeta_1, \ldots, \zeta_n)$ is a common root of the polynomials $\hat{f}^*, f_1^*, \ldots, f_n^*$ in the variables $y_0, y_1, \ldots, y_n$. As a consequent, it is also a common root of the ideal $\mathcal{I} := \langle \mathcal{F}^* \cup \{\hat{f}^*\} \rangle$. Now, let $\preceq$ be an arbitrary elimination term ordering on $\mathcal{T}$ for $y_1, \ldots, y_n$. Since a Gröbner basis is just a specific representation of an ideal, making transition to $\mathcal{GB}_{\preceq}(\mathcal{I})$ does not affect the common roots of $\mathcal{I}$. Finally, $\mathcal{GB}_{\preceq}(\mathcal{I}) \cap \mathbb{Q}[y_0]$ is a subset of $\mathcal{GB}_{\preceq}(\mathcal{I})$ and does not mention any of $y_1, \ldots, y_n$. Furthermore, it is a representation of the elimination ideal $\mathcal{I} \cap \mathbb{Q}[y_0]$, and thus, x is a common root of $\mathcal{I} \cap \mathbb{Q}[y_0]$. $\square$

6 Fraud detection in enterprises

We apply the presented methodology to an example in the field of auditing. During an audit, the auditor has to estimate the risk if the balance sheet has been manipulated premeditatedly. It depends on the outcome of this risk estimation whether the audit will be done more intensely. The estimation typically results from investigating risk indicators, so-called *red flags* (cf. Tab. 1). A collocation of appropriate red flags has emanated from a study of Albrecht and Romney (Albrecht and Romney 1986) and is discussed in (Terlinde 2003). The authors presented a list of possible fraud indicators to auditors who uncovered balance sheet manipulation and asked them which indicators had been relevant. A comparison group of auditors specified their observed indicators in cases where no balance sheet manipulation was detected. An excerpt of the red flags categorized as significant is shown in Tab. 2. The last two columns indicate the relative frequency of the mentioned red flag in dependence of wether balance sheet manipulation is present or not.

Red flag	Description
R_1	Corporate officers have personal high debts or losses
R_2	Corporate officers are greedy
R_3	Close connections between corporate officers and distributors
R_4	Lack of established and consistent rules for the employees
R_5	Doubts regarding the integrity of the corporate officers
R_6	Inappropriate complex corporate structure
R_7	Company Management is under high pressure to present positive operating profit
R_8	High amount of unusual transactions at the end of the accounting year
R_9	Unfair payment practices
R_{10}	Decreasing profit qualities
R_{11}	Business with affiliated companies
R_{12}	Required internal revision is missing

Table 1: Description of the red flags

Red flag	Balance sheet manipulated	Not manipulated
R_1	$0.44 = 11/25$	$0.05 = 1/20$
R_2	$.41 = 41/100$	$.06 = 3/50$
R_3	$.48 = 12/25$	$.10 = 1/10$
R_4	$.55 = 11/20$	$.11 = 11/100$
R_5	$.59 = 59/100$	$.19 = 19/100$
R_6	$.36 = 9/25$	$.08 = 2/25$
R_7	$.31 = 31/100$	$.06 = 3/50$
R_8	$.40 = 2/5$	$.11 = 11/100$
R_9	$.24 = 6/25$	$.03 = 3/100$
R_{10}	$.27 = 27/100$	$.06 = 3/50$
R_{11}	$.50 = 1/2$	$.23 = 23/100$
R_{12}	$.40 = 2/5$	$.13 = 13/100$

Table 2: Red flags that are significant for balance sheet audit

The data from Tab. 2 may now serve as a knowledge base to which a query is made (cf. (Finthammer, Kern-Isberner, and Ritterskamp 2007)). Therefore, the auditor takes a note of which red flags apply to the inspected company. In a more general case with any number of red flags $R_1, \ldots, R_n$, the knowledge base looks like

$$\begin{aligned} \mathcal{KB}_{audit} := \{&(R_1|M)[x_1], \ldots, (R_n|M)[x_n], \\ &(R_1|\overline{M})[x_{n+1}], \ldots, (R_n|\overline{M})[x_{2n}]\} \end{aligned} \tag{9}$$

where M stands for "case of balance sheet manipulation"

($\overline{M}$ stands for "case of no balance sheet manipulation", respectively) and $x_1, \ldots, x_{2n}$ denote the relative frequencies of occurrence of the corresponding red flags. A typical inference query would be to ask for the probability of the presence of a balance sheet manipulation given the available information on an enterprise, i.e.,

$$\mathcal{KB}_{audit} \mid\!\sim_{\mathcal{ME}} (M|O)[x] \,? \tag{10}$$

based on the observation

$$O := \left(\bigwedge_{1 \le i \le s} R_{m_i} \right) \wedge \left(\bigwedge_{s+1 \le i \le r} \overline{R}_{m_i} \right) \tag{11}$$

with $m_i \in \{1, \ldots, n\}$ for $1 \le i \le r$ and $m_i \ne m_j$ for $i \ne j$. This means that the red flags $R_{m_1}, \ldots, R_{m_s}$ apply to the inspected company and the red flags $R_{m_{s+1}}, \ldots, R_{m_r}$ do not. Furthermore, no information about the remaining red flags is available. Note that this ignorance on the remaining red flags is not possible with Bayesian networks and, in addition, that MaxEnt does not need a prior probability of M.

To give an answer to the query (10), we apply the methodology presented above. Therefore, we have to determine the polynomial analogue to $\mathcal{KB}_{audit}$. From (4), we get

$$\begin{aligned} f_{i,audit} := (q_i - p_i)\, y_i^{q_i} & \sum_{\omega \models MR_i} \prod_{\substack{j \ne i \\ \omega \models MR_j}} y_j^{q_j} \prod_{k=n+1}^{2n} y_k^{p_k} \\ & - p_i \sum_{\omega \models M\overline{R_i}} \prod_{\substack{j \ne i \\ \omega \models MR_j}} y_j^{q_j} \prod_{k=n+1}^{2n} y_k^{p_k} \end{aligned}$$

for $1 \le i \le n$. The polynomials corresponding to the conditionals $(R_1|\overline{M})[x_{n+1}], \ldots, (R_n|\overline{M})[x_{2n}]$ in $\mathcal{KB}_{audit}$ look similarly, so we focus on the polynomials $f_{1,audit}, \ldots, f_{n,audit}$ in the following. First, we want to simplify the polynomial expressions. As $\sum_{\omega \models M\dot{R}_i} \prod_{\substack{j \ne i \\ \omega \models MR_j}} y_j^{q_j}$ is just the sum of each combination of the terms $y_j^{q_j}$ for $1 \le j \le n$ with $j \ne i$ in both cases, i.e., $\dot{R}_i = R_i$ or $\dot{R}_i = \overline{R}_i$, we introduce vectors $\varepsilon^i := (\varepsilon_1^i, \ldots, \varepsilon_n^i) \in \{0,1\}^n$ with $\varepsilon_i^i = 0$. The other entries may be 0 or 1 where the case $\varepsilon_j^i = 1$ matches the condition $\omega \models MR_j$, and we sum up over every possible combination of them. This leads to

$$\begin{aligned} f_{i,audit} = (q_i - p_i)\, y_i^{q_i} & \sum_{\varepsilon^i} \prod_{j \ne i} (y_j^{q_j})^{\varepsilon_j^i} \prod_{k=n+1}^{2n} y_k^{p_k} \\ & - p_i \sum_{\varepsilon^i} \prod_{j \ne i} (y_j^{q_j})^{\varepsilon_j^i} \prod_{k=n+1}^{2n} y_k^{p_k} \end{aligned} \tag{12}$$

for $1 \le i \le n$. Note that ε_i^i does not appear in (12) and is only set vacuously to 0 so that the sums in (12) are not exploited twice. Now, it is obvious that both sums appearing in (12) are exactly the same. Therefore,

$$\gcd(f_{i,audit}^+, f_{i,audit}^-) = \sum_{\varepsilon^i} \prod_{j \ne i} (y_j^{q_j})^{\varepsilon_j^i} \prod_{k=n+1}^{2n} y_k^{p_k}$$

holds, and it follows that $f_{i,audit}^* := (q_i - p_i)\, y_i^{q_i} - p_i$ for $1 \le i \le n$. As the same argumentation may be given for the remaining polynomials $f_{n+1,audit}, \ldots, f_{2n,audit}$, we get

$$f_{i,audit}^* = (q_i - p_i)\, y_i^{q_i} - p_i \tag{13}$$

for $1 \le i \le 2n$, i.e., for every polynomial counterpart of the conditionals in $\mathcal{KB}_{audit}$. This result demonstrates the power of the simplification step (5) in full clarity and reflects the symmetric inner structure of the knowledge base $\mathcal{KB}_{audit}$. As we are interested in positive real roots,

$$y_i = \left(\frac{p_i}{q_i - p_i} \right)^{1/q_i}$$

holds for $1 \le i \le 2n$, and the effects of the conditionals in $\mathcal{KB}_{audit}$ are given as $\alpha_i = \frac{p_i}{q_i - p_i}$ by reversing the substitution (3). In order to answer the inference query (10), it is necessary to derive the polynomial $\hat{f}_{audit}^*$ which is (8) applied to $\mathcal{KB}_{audit}$. This can be done by performing analogous steps as before. With the use of the auxiliary set $\mathcal{M} := \{m_1, \ldots, m_s\}$ (cf. (11)) and the vectors $\varepsilon := (\varepsilon_1, \ldots, \varepsilon_n) \in \{0,1\}^n$ with $\varepsilon_{m_i} = 0$ for $1 \le i \le r$ (again, the remaining entries may be 0 or 1, and we sum up over every combination), we get

$$\begin{aligned} \hat{f}_{audit}^* := y_0 \Bigg(& \prod_{\substack{1 \le j \le n \\ j \in \mathcal{M}}} y_j^{q_j - p_j} \prod_{\substack{1 \le k \le n \\ k \notin \mathcal{M}}} y_{k+n}^{p_{k+n}} \sum_{\varepsilon} \prod_{i=1}^{n} (y_i^{q_i})^{\varepsilon_i} \\ & + \prod_{\substack{1 \le j \le n \\ j \in \mathcal{M}}} y_{j+n}^{q_{j+n} - p_{j+n}} \prod_{\substack{1 \le k \le n \\ k \notin \mathcal{M}}} y_k^{p_k} \sum_{\varepsilon} \prod_{i=1}^{n} (y_{i+n}^{q_{i+n}})^{\varepsilon_i} \Bigg) \\ & - \prod_{\substack{1 \le j \le n \\ j \in \mathcal{M}}} y_j^{q_j - p_j} \prod_{\substack{1 \le k \le n \\ k \notin \mathcal{M}}} y_{k+n}^{p_{k+n}} \sum_{\varepsilon} \prod_{i=1}^{n} (y_i^{q_i})^{\varepsilon_i}. \end{aligned} \tag{14}$$

Example 1. *Assume that an auditor examines a balance sheet with the help of a checklist based on the red flags given in Tab. 2. Then, his background knowledge is*

$$\begin{aligned} \mathcal{KB}_{audit} = \{ & (R_1|M)[x_1], \ldots, (R_{12}|M)[x_{12}], \\ & (R_1|\overline{M})[x_{13}], \ldots, (R_{12}|\overline{M})[x_{24}] \} \end{aligned}$$

with the probabilities (cf. Tab. 2)

$$\begin{array}{llll} x_1 = 11/25, & x_2 = 41/100, & x_3 = 12/25, & x_4 = 11/20, \\ x_5 = 59/100, & x_6 = 9/25, & x_7 = 31/100, & x_8 = 2/5, \\ x_9 = 6/25, & x_{10} = 27/100, & x_{11} = 1/2, & x_{12} = 2/5, \\ x_{13} = 1/20, & x_{14} = 3/50, & x_{15} = 1/10, & x_{16} = 11/100, \\ x_{17} = 19/100, & x_{18} = 2/25, & x_{19} = 3/50, & x_{20} = 11/100, \\ x_{21} = 3/100, & x_{22} = 3/50, & x_{23} = 23/100, & x_{24} = 13/100. \end{array}$$

The simplified polynomials (13) relating to $\mathcal{KB}_{audit}$ are

$$\begin{array}{ll} f_{1,audit}^* = 14\, y_1^{25} - 11, & f_{2,audit}^* = 59\, y_2^{100} - 41, \\ f_{3,audit}^* = 13\, y_3^{25} - 12, & f_{4,audit}^* = 9\, y_4^{20} - 11, \\ f_{5,audit}^* = 41\, y_5^{100} - 59, & f_{6,audit}^* = 16\, y_6^{25} - 9, \\ \ldots & \end{array}$$

Note that the example is only restricted to twelve red flags in order to make the formulas more readable, and note that the equations $f^*_{i,audit} = 0$ can be solved easily.

Furthermore, we assume that during the audit, evidence for the presence of R_1, R_3, R_4, R_9, R_{12} and for the absence of R_5, R_6, R_7, R_8, R_{10} is found, while nothing is known about the other red flags. The proper inference query is

$$\mathcal{KB}_{audit} \mid\!\sim_{\mathcal{ME}} (M|O)[x]\ ?$$

with $O = R_1\ R_3\ R_4\ \overline{R_5}\ \overline{R_6}\ \overline{R_7}\ \overline{R_8}\ R_9\ \overline{R_{10}}\ R_{12}$. The resulting polynomial according to (14) is

$$\begin{aligned}\hat{f}^*_{audit} = y_0\, (&y_1^{14}\, y_3^{13}\, y_4^{9}\, y_9^{19}\, y_{12}^{3}\, y_{14}^{3}\, y_{17}^{19}\, y_{18}^{2}\, y_{19}^{3}\, y_{20}^{11}\, y_{22}^{3}\, y_{23}^{23} \\ &(1 + y_2^{100} + y_{11}^{2} + y_2^{100}\, y_{11}^{2}) + y_{13}^{19}\, y_{15}^{9}\, y_{16}^{89}\, y_{21}^{97}\, y_{24}^{87} \\ &y_2^{41}\, y_5^{59}\, y_6^{9}\, y_7^{31}\, y_8^{2}\, y_{10}^{27}\, y_{11}\, (1 + y_{14}^{50} + y_{23}^{100} + y_{14}^{50} \\ &y_{23}^{100})) - y_1^{14}\, y_3^{13}\, y_4^{9}\, y_9^{19}\, y_{12}^{3}\, y_{14}^{3}\, y_{17}^{19}\, y_{18}^{2}\, y_{19}^{3} \\ &y_{20}^{11}\, y_{22}^{3}\, y_{23}^{23}\, (1 + y_2^{100} + y_{11}^{2} + y_2^{100}\, y_{11}^{2}).\end{aligned}$$

Now, it is not very difficult to find x from determining the y_0-component of a common root of $\hat{f}^*_{audit}$ and $f^*_{i,audit}$ according to Theorem 2. The auditor has to assume balance sheet manipulation with a probability of $x \approx 99,998\%$ which is so high because of the great amount of observed red flags.

7 Conclusion

In this paper, we presented a novel approach to calculating probability distributions according to the MaxEnt principle by means of computer algebra. We explored the algebraic inner structure of such distributions (Kern-Isberner 2001) which implements conditional-logical features to encode the information given by a probabilistic knowledge base by way of a system of polynomial equations. Any solution of this equation system defines the MaxEnt distribution appertaining to the knowledge base; in fact, the MaxEnt distribution is determined uniquely by the system (cf. (Kern-Isberner 2001)). Our approach allows an algebraic understanding of MaxEnt inferences by means of Gröbner bases theory and symbolic computation and thus connects profound mathematical methods with inductive probabilistic reasoning on maximum entropy, a principle that has often been perceived as a black box methodology. To our knowledge, our approach is the first to make this connection explicit. Previous work by Dukkipati (cf. (Dukkipati 2009)) investigated a similar link between Gröbner bases and MaxEnt distributions in the field of statistics but does not address any issues of knowledge representation, in particular, Dukkipati's approach does not consider knowledge bases nor inferences.

The approach presented in this paper also makes it possible to perform generic MaxEnt inferences, i.e., yielding symbolic inferred MaxEnt probabilities without knowing the given probabilities of the knowledge base explicitly (cf. (Kern-Isberner, Wilhelm, and Beierle 2014)). As part of our current and future work, we continue the elaboration of computational symbolic reasoning according to the MaxEnt principle in order to improve the understanding of MaxEnt reasoning and make it more transparent and usable for applications.

References

Albrecht, S. W., and Romney, M. B. 1986. Red-flagging management fraud: A validation. *Advances in Accounting* 3:323–333.

Buchberger, B. 2006. An algorithm for finding the basis elements in the residue class ring modulo a zero dimensional polynomial ideal. *J. of Symbolic Comput.* 41(3-4):475–511.

Cowell, R.; Dawid, A.; Lauritzen, S.; and Spiegelhalter, D. 1999. *Probabilistic networks and expert systems*. New York Berlin Heidelberg: Springer.

Cox, D. A.; Little, J.; and O'Shea, D. 2007. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer.

Dukkipati, A. 2009. Embedding maximum entropy models in algebraic varieties by Gröbner bases methods. In *ISIT*, 1904–1908. IEEE.

Finthammer, M.; Kern-Isberner, G.; and Ritterskamp, M. 2007. Resolving inconsistencies in probabilistic knowledge bases. In *Proceedings of the 30th Annual German Conference on AI, KI 2007*, volume 4667 of *LNCS*, 114–128. Springer.

Gärdenfors, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, Mass.: MIT Press.

Jaynes, E. 1983. Where do we stand on maximum entropy? In *Papers on Probability, Statistics and Statistical Physics*. Dordrecht, Holland: D. Reidel Publishing Company. 210–314.

Kern-Isberner, G.; Wilhelm, M.; and Beierle, C. 2014. Probabilistic knowledge representation using Gröbner basis theory. In *International Symposium on Artificial Intelligence and Mathematics (ISAIM 2014), Fort Lauderdale, FL, USA, January 6-8, 2014*.

Kern-Isberner, G. 1998. A note on conditional logics and entropy. *International Journal of Approximate Reasoning* 19:231–246.

Kern-Isberner, G. 2001. *Conditionals in nonmonotonic reasoning and belief revision*. Springer, Lecture Notes in Artificial Intelligence LNAI 2087.

Paris, J. 1994. *The uncertain reasoner's companion – A mathematical perspective*. Cambridge University Press.

Paris, J. 1999. Common sense and maximum entropy. *Synthese* 117:75–93.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. San Mateo, Ca.: Morgan Kaufmann.

Rödder, W., and Meyer, C. 1996. Coherent knowledge processing at maximum entropy by SPIRIT. In *Proceedings Uncertainty in AI, UAI 1996*, 470–476.

Shore, J., and Johnson, R. 1980. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* IT-26:26–37.

Terlinde, C. H. 2003. *Aufdeckung von Bilanzmanipulationen im Rahmen der Jahresabschlussprüfung – Ergebnisse theoretischer und empirischer Analyse*. PhD dissertation, Universität Duisburg-Essen.