Sapienza Università di Roma

Dottorato di Ricerca in Statistica Metodologica

Tesi di Dottorato XXX Ciclo – anno 2016-2017

Dipartimento di Statistica, Probabilità e Statistiche Applicate

Generalized Belief Change

with Imprecise Probabilities and Graphical Models

Sabina Marchetti

Alle mie ragazze, e a L.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to Professor Piero Manfredi and Professor Gianpaolo Scalia Tomba, for giving me the most valuable experience of taking part to prestigious research projects, while doing my PhD research. I am deeply thankful for their kindness and trust. Our collaboration allowed me to be part of the following research projects:

- FIRB Research Project: *Strategie per la prevenzione del carcinoma della cervice uterina: prospettive a medio e lungo termine in Italia*, of Dep. Health Sciences at Università degli Studi di Firenze; see [172]. Some methods and results from this research are reported in the Appendix.

- Norwegian Institute of Public Health (NIPH)'s Research Project: *Burden of varicella-zoster virus infection in Norway*, of Dep. Economics and Management, Università di Pisa; see [173].

I seize the opportunity to extend my thanks to Dr. Luca Faustini and Dr. Giorgio Guzzetta, for their valuable professional support and friendship.
My PhD scholarship was funded by Università Europea of Rome, whom I am thankful for.
Research for conference paper [171] was carried out during my visiting period at Istituto Dalle Molle di Studi sull'Intelligenza Artificiale in Manno (Lugano, Switzerland), under the supervision Dr. Alessandro Antonucci. I would like to offer him my special thanks for his kindness and valuable suggestions, and for opening my research to the enchanting topic of imprecise probabilities.
I am also deeply thankful to Prof. Paola Vicard (Roma Tre University) for her trust and support.

Finally, I am grateful to both external evaluators, whose valuable comments and proposed suggestions were much appreciated, and carefully considered in the updated version of this manuscript.

Contents

List of Figures

List of Tables

Abbreviations

AGM Alchourrón-Gärdenfors-Makinson (Theory or Postulates or Axioms)

AR Adjustment Rule

BN Bayesian Network

CAR Coarsening At Random

CC Cervical Cancer

CCM Cano-Cano-Moral (Transformation)

CCoPK Credal Conditional Probability Kinematics

CCPT Credal Conditional Probability Table

CD Chan-Darwiche (Distance)

CE Causal Irrelevance

CH Convex Hull

CIN Cervical Intraepithelial Neoplasia

CIR Conservative Inference Rule

CIS Cancer In Situ

CN Credal Network

CoPK Conditional Probability Kinematics

CoSE Conditional Soft Evidence

CPC Conservative PC (Algorithm). PC are the initials of Peter (Spirtes) and Clark (Glymour) who first proposed the algorithm

CPK Credal Probability Kinematics

CPLDAG Completed Partially Labeled Acyclic Directed Graph

CPT Conditional Probability Table

CS Credal Set

CS-WI Context-Specific Wagner Independence

CSE Credal Soft Evidence

CUR Conservative Updating Rule

CVE Credal Virtual Evidence

DAG Acyclic Directed Graph

DRC Dempster's Rule of Combination

ECE Elementwise Causal Irrelevance

ECPT Extensive (Credal) Conditional Probability Table

EI Epistemic Independence

EM Expectation Maximization (Algorithm)

ER Epistemic Irrelevance

GBR Generalized Bayes Rule

GEMA Generalized Expectation Maximization Algorithm

HPV Human Papilloma Virus

ICK Imaginary Conditional Kinematics

IK Imaginary Kinematics

iLogOp Informed LogOp (Logarithmic Pooling Operator)

IPFP Iterative Proportional Fitting Procedure

JT Junction Tree

KL Kullback-Liebler (Divergence or -Projection)

KM Katzuno-Mendelzon (Postulates or Axioms)

LDAG Labeled Acyclic Directed Graph

LHS Latin Hypercube Sampling

MA Multiple Adams Conditioning

MAP Maximum A Posteriori

MAR Missing At Random

MBR Multiple Belief Revision

MBS Multivalued Brier Score

MPE Most Probable Explanation

NBC Naïve Bayes Classifier

NPO Naïve Pooling Operator

nSS Nonseparately Specified

ODE Ordinary Differential Equation

PGM Probabilistic Graphical Model

PK Probability Kinematics

PMF Probability Mass Function

PO Pooling Operator

r.v. Random variable

RS Reliability Score

SE Soft Evidence

SI Stochastic Independence

SIR Susceptible-Infected-Removed (Model)

SL Structural Learning

SoP System of Play

SS Separately Specified

St_1 Strong Independence on Distribution

St_2 Strong Independence on Decomposition

TAN Tree Augmented Naïve Bayes Classifier

UCU Uncertain Credal Updating

VE Virtual Evidence

VZV Varicella Zoster Virus

WI Wagner Independence

Introduction

This thesis is concerned with probabilistic reasoning, under generalized conditions of uncertainty and imprecision. It provides a theoretical investigation of probabilistic belief revision, successively intended as evidence propagation in statistical multivariate models.

Ch. 1 introduces fundamental concepts from the literature of imprecise probabilities, with a major focus on the issue of independence and fulfillment of graphoid axioms (Sec. 1.2). Probabilistic graphical models are successively introduced in Sec. 1.3. Particularly those based on an acyclic directed graphical component are considered, namely Bayesian networks (Sec. 1.3.1) and their credal counterpart (Sec. 1.3.2).

Ch. 2 gives an overview of the general subject of belief change theory, also known as AGM theory. This studies the way a doxastic agent adjusts her set of beliefs upon new information, in a static context. Sec. 2.1 eventually specializes the discussion to what is referred throughout as *probabilistic belief revision*, dealing with probability mass functions. There, general properties of functionals for belief adjustment are introduced, and *probability kinematics* [138] are accounted for as desirable mechanics for adjusting an agent's belief upon information, in the standard case of *soft* (or probabilistic) evidence [244] (Sec. 2.2.1). The rationale behind this choice derives from a *minimal change* principle, advocated by several authors in the literature. Among others, Rott's definition of belief change, as rational integration of new pieces of information into a doxastic agent's knowledge structure [215], is worth mentioning. There, new information is given the informal definition of a *structure realized in the physical world, that is suitable to be interpreted or exploited by some receiver in a reasonable way*[1] [215, Sec.2.1]. Following Rott, the framework considered in the present work is synthesized by the following dynamics: given some piece of information I,

$$\text{Prior Belief} \longrightarrow_I \text{Posterior Belief}$$

[1]What Rott means by *reasonable way* is behavioral success of the agent.

Roughly, new information acts as an input to an agent's belief state. Absorption of information by an agent's system of belief ought not be altered in its form, based on what will be introduced as a *conservativeness* (or *rigidity*) principle. This latter requires invariance of the information's relevance on the whole set of beliefs.
Based on the well-known Jeffrey's rule for belief revision with probabilistic evidence, several extensions to the former are proposed throughout the chapter, increasingly accounting for i) conditional (contex-specific) knowledge (Sec. 2.2.2), ii) imprecision (Sec. 2.2.3), and iii) violations of the partiality principle (Sec. 2.3), properly characterized. Axioms for extended probability kinematics are introduced accordingly, as well as theoretical results on the properties of the functionals that are proposed.

The discussion is further specialized to reasoning with probabilistic graphical models in Ch. 3. Most results are derived from the relevant work of Chan and Darwiche [35], that bridged the, deeply distinct in their meaning, concepts of virtual [194] and soft evidence, by showing their inter-reducibility in the standard probabilistic framework. Contributions from Sec. 3.2 extend Pearl's method for virtual evidence absorption by a Bayesian network to the case of non-standard evidence (considered by [35]), including context-specific instances and credal information. Indeed, propagation of credal (uncertain) evidence renders the whole model imprecise, i.e. it produces a Credal network. Sec. 3.3 moves a step forward and accounts for the case of uncertain belief propagation in a Credal network. Due to the computational complexity of the inferential task, approximate techniques are proposed to deal with such a situation.

In Ch. 4 probabilistic belief change is extended to belief aggregation, or opinion pooling. This is intended as a natural extension of classical (probabilistic) belief revision to the case of multiple overlapping sources of information. Analogously to previous chapters, the discussion considers generalized settings, and hence properly defined functionals (Sec. 4.2) and their properties (previously introduced in Sec. 4.1), with a focus on their implementation with probabilistic graphical models.

Finally, Ch. 5 sheds a light on the iterated case, when probabilistic information is available on a subset of random variables. Basic concepts and principles are introduced, as well as a systematization of the setup: as either accounting for an priority (or ordering) applicable on the information or not; this latter case is characterized as *multiple* (or *simultaneous*) belief revision, as opposed to iterated. Among others, the commutativity principle is considered in Sec. 5.1, where results on the conditions that guarantees it are provided for revision processes involving sharp probability distributions only. Both iterated and multiple belief revision are eventually considered in their graphical implementation (Sec. 5.3).

Contributions from this work may also be found in conference papers [171], 31th FLAIRS (Florida Artificial Intelligence Research Society) Conference, and [170], to be presented at UAI (Uncertainty in Artificial Intelligence) 2018 Conference. Joint work with A. Antonucci [171], was nominated for best conference paper. Finally, contributions from the Appendix shall also be found in [173, 172].

Chapter 1

Probabilistic Reasoning Under Uncertainty and Imprecision

1.1 Modeling Uncertainty

Probability theory addresses various distinct concepts in the statistical, mathematical and philosophical literature. Coarsely, it may be understood in the *objective* or *subjective* traditions. The first, also referred to as classical probability, dates back to the seminal work of De Moivre in 1718 [80], and was successively given a systematic discussion and axiomatization by, among others, [151]. It is based on frequentist theory and relies on the physical repeatability of an experiment to capture variability, or randomness, of its outcomes in the limit. i.e. the limiting frequency of an outcome reproduces its true probability. Let You denote any doxastic subject, holding coherent probabilistic assessments [79, 257]. In the words of Smets [232], objective probabilities "exist outside of You", as a property of the world.[1] Subjectivist approaches to probability address such questions by considering the *epistemic* beliefs of an agent - or You. They date back to Bayes' posthumous seminal work of 1763 [20] and, lately, among others, to [208, 79]. Epistemic belief and uncertainty are represented by formal probability distributions intended as *bets* on events. In such a way, when facing a decision Your behavior , as well as Your betting disposition, are modelled: probabilities serve as prices You are willing to pay to enter a game that pays back one unit of money if the event occurs, 0 otherwise. Such probabilities are used for knowledge representation, modeling personal beliefs (and their strength) and doing rational reasoning.

[1] Several objections may be raised against objectivist concepts of probability, starting from the non-repeatability of some events. Furthermore, does chance exist if Your uncertainty about reality is not introduced? Also, is the world outside itself deterministic? [232]

Let (Ω, Σ, P) be any discrete *probability space*. To our purposes, Ω denotes the (countable) *sample space* of all mutually exclusive and exhaustive sample points ω, while Σ is the σ-field induced by Ω. A collection Σ of subsets of Σ is called an algebra, or a field, over Ω if the following conditions are met:

1. $\Omega \in \Sigma$,
2. If $\alpha \in \Sigma$, then $\neg\alpha \in \Sigma$, where $\neg$ represents negation,
3. If $\alpha, \beta \in \Sigma$, then $\alpha \cup \beta \in \Sigma$.

Σ is called a σ-field when 3. extends to countable union.
Let $|\cdot|$ denote cardinality of its argument, Σ is the collection of all $2^{|\Omega|}$ subsets of Ω, called *events*. By the *Kolmogorov axioms*, P is probability measure on Σ, defining the mapping $\Sigma \to [0,1] = \Delta$ (probability simplex) if it holds:

1. $P(\Omega) = \sum_{\omega\in\Omega} P(\omega) = 1$, (Normalization)
2. $P(\alpha) = \sum_{\omega\in\alpha} P(\omega) \geq 0$, $\alpha \in \Sigma$, (Non-Negativity)
3. $P\left(\cup_{j=1}^{\infty}\alpha_j\right) = \sum_{j=1}^{\infty} P(\alpha_j)$, for any countable collection of mutually disjoint events in Σ. (σ-Additivity)

As a remark, Kolmogorov axioms apply to both objective and subjective concepts of probability.[2] Let X denote any random variable (r.v.) taking all possible values in Ω_X, set of mutually exclusive and exhaustive *potential outcomes* x. We interchangeably refer to the elements of Ω_X as states, realizations or instantiations throughout. X is called a *discrete* r.v. if $|\Omega_X| < \infty$, and its behavior may be described by a probability mass function (PMF) P_X, mapping each element $x \in \Omega_X$ to a real value in the probability simplex. P_X is *continuous* whenever Ω_X is uncountably infinite. Any *absolutely continuous* r.v. is described by a probability density function.

Let $P_{\mathbf{V}}$ be a strictly positive PMF over joint r.v. $\mathbf{V} = \{X_0, \ldots, X_n\}$, $n \geq 0$. We use bold letters to denote sets. $\mathbf{V}$ is a joint random variable whose generic realization is $\mathbf{v} \in \Omega_{\mathbf{V}} \subseteq \times_{X\in\mathbf{V}}\Omega_X$, where $\times$ denotes Cartesian product of every X_i's sample space, for each $X \in \mathbf{V}$. It is easy to see $\Omega \equiv \Omega_{\mathbf{V}}$ and $\omega \equiv \mathbf{v} = (x_0, \ldots, x_n)$. We shall use Ω and $\Omega_{\mathbf{V}}$ interchangeably throughout. Also, we write the joint PMF as $P_{\mathbf{V}}(X_0 = x_0, \ldots, X_n = x_n) = P(x_0, \ldots, x_n)$, to avoid cumbersome notation.
Let $\mathbf{V} = \{X, Y\}$, the probability of event $(X = x)$ conditional on $(Y = y)$ is given

[2]See [208, 212, 79, 243]. See also [24, 125] on departures from axioms 1-3.

by[3]:
$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}\,. \tag{1.1}$$
We refer to Eq. (1.1) as *conditioning*, or *updating*.[4] By conditioning, P is projected from Ω toward a single element of the (coarse) partition that Y induces on it. This way, $P(x|x) = 1$. When Eq. (1.1) is used to update P upon learned information, successive (e.g. in time) learning of any other state x' from Ω_X is undefined (or yields inconsistencies).
As a remark, let us stress proposition "$(X, Y) = (x, y)$ is possible" does not imply, nor it is implied, by "$(X = x)$ and $(Y = y)$ are separately possible". X and Y are called *logically independent* conditional on a third r.v. Z if event (x, y, z) is possible whenever both (x, z) and (y, z) are possible. If this is the case, for any fixed $z \in \Omega_Z$, P is defined on the product space $\Omega_X \times \Omega_Y$. We shall now provide basic definitions for independence concepts.

Definition 1 (Stochastic Independence)**.** *Let* $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ *be any mutually disjoint sets of random variables.* $\mathbf{X}$ *is stochastically independent (SI) of* $\mathbf{Y}$ *conditional on* $\mathbf{Z}$ *iff it holds:*
$$P(\mathbf{x}, \mathbf{y}|\mathbf{z}) = P(\mathbf{x}|\mathbf{z})P(\mathbf{y}|\mathbf{z})$$
for each $\mathbf{x} \in \Omega_{\mathbf{X}}$,$\mathbf{y} \in \Omega_{\mathbf{Y}}$ *and* $\mathbf{z} \in \Omega_{\mathbf{Z}}$*, provided* P *is strictly positive over* $\Omega_{\mathbf{Z}}$*. When SI holds, we write* $I(\mathbf{X}, \mathbf{Y}; \mathbf{Z})$*, otherwise* $\neg I(\mathbf{X}, \mathbf{Y}; \mathbf{Z})$*.*

Def. 1 trivially applies to unitary sets of random variables. Also, $P(\mathbf{x}, \mathbf{y}) = P(\mathbf{x})P(\mathbf{y})$ if $\mathbf{z} = \emptyset$.

Definition 2 (Context-Specific Independence [26])**.** *Random variables* X *and* Y *are context-specific independent (CSI) conditional on* $\mathbf{Z} = \mathbf{z}^*$ *iff it holds:*
$$I(X, Y; \mathbf{z}^*) \text{ whereas } \neg I(X, Y; \mathbf{Z})$$
for some $\mathbf{z}^* \in \Omega_Z$*.*

CSI reduces to SI whenever it applies to all elements of $\Omega_{\mathbf{Z}}$. Context-specific extensions of SI were also considered by [112, 107]. A further concept, of *contextual weak independence*, was introduced by [265], allowing irrelevance relationships to hold within elements of refined partitions of a given joint sample space. Analogously to CSI, contextual weak independence reduces to weak independence whenever it applies to all conditioning states; this is in turn more general than SI. We avoid

[3]Remember we assumed P to be strictly positive. In the general case, it must be checked $P(y) > 0$.
[4]Eq. (1.1) is usually referred to as *Bayes* rule in the PGM's and imprecise literature.

formal introduction of these mentioned concepts, as they are outside the scope of the present work. We will return to CSI in Sec. 1.3[5].
As previously argued, it might be the case events (x, z) and (y, z) are possible, while (x, y, z) is not. If this is the case we usually set $P(x, y|z) = 0$ [54]. Based on [257, Sec. 9.1], logical independence of X and Y conditional on Z shall be a prerequisite for the definition of SI. Such strengthening applies to zero-probability event $P(x|z)$ or, either, $P(y|z)$, for some (x, z) or (y, z), respectively.[6]
We shall now introduce graphoid axioms, that elegantly connect probability to graph theory. A graphoid shall be intended as a collection of statements over sets of random variables, aimed to graphically, i.e. via paths on graphs, represent the abstract properties [54] that ought to be satisfied by a given concept for irrelevance. In our setup, irrelevance is given a probabilistic interpretation.

Definition 3 (Graphoid Axioms [195, 67, 241, 54])**.** *Let $*$ denote any conditional irrelevance relation among X and Y, conditional on Z; formally, $(X * Y|Z)$ reads "X is irrelevant to Y when Z is known". $*$ is called a graphoid if it satisfies the following:*

1. $X * Y|Z$ *if and only if* ($\iff$) $Y * X|Z$, *(Symmetry)*
2. $X * Y|X$, *(Redundancy)*
3. $X * (Y \cup W)|Z$ *implies* ($\implies$) $X * Y|Z$, *(Decomposition)*
4. $X * (Y \cup W)|Z \implies X * Y|(Z \cup W)$, *(Weak Union)*
5. $X * Y|Z \wedge X * W|(Y \cup Z) \implies X * (Y \cup W)|Z$, *(Contraction)*
6. *If Y and W are logically independent conditional on Z, then* $X * Y|(W \cup Z) \wedge X * W|(Y \cup Z) \implies X * (Y \cup W)|Z$. *(Intersection)*

A relation $*$ that satisfies conditions 1) to 5) is called a semi-graphoid, while an a-graphoid[7] satisfies conditions 2) to 6).

SI is a graphoid, whenever the P is strictly positive; in such a case, $X * Y|Z$ shall be equivalently written as $I(X, Y; Z)$.

Reasoning in fully deterministic settings is modeled by *logic*. When uncertainty is introduced, logic extends to probability theory. In the upcoming section, uncertain reasoning is further extended to imprecise probabilities [257].

[5] The interested reader shall refer to [274] on inference with CSI.

[6] See [91] on full conditional measures, and [42, 43, 246, 248, 56] on independence concepts and related properties.

[7] See [247] on a-graphoids and L-separation.

1.2 Modeling Imprecision

The general notion of imprecise probabilities refers to mathematical models that relax the rationality requirements of classical probability theory [36] by allowing probabilities not to take sharp, or *precise*, numerical values [54]. They combine probability theory with logic (e.g. propositional [135], first-order [19]) and are used in a wide variety of fields, ranging from statistics to artificial intelligence, psychology and economics. Although several distinct theories exist for imprecise probabilities, we take Walley's coherent lower previsions [257] as reference.
As a physical - objective - justification, while models built under uncertainty are induced by complete records of data, those incomplete[8] yield imprecision [4], and a true parameter is assessed to vary within a range, defined by upper and lower probability values. Imprecision may arise in the subjectivist framework when highly incomplete or conflicting information [8] is gathered to any rational agent, whose belief may be affected by lack of introspection [183]. It may also result from indeterminacy [257], as we shall discuss below. Also, imprecise theory accounts for a qualitative approach to knowledge representation, incorporating the subjectivist point of view, and serve as the basis for the theory of desirable gambles (see [257]).

Lower and upper previsions use *gambles* to represent behavioral beliefs. A gamble corresponds to a transaction (or a decision) yielding rewards, usually referred to as *utilities*, when states of the world occur [15]. Let Ω be the set of all possible outcomes, a gamble f defines a bounded mapping from Ω to $\mathbb{R}$. f models Your uncertainty about incoming events $\omega \in \Omega$. Once a gamble is accepted, rewards are expressed on a linear (to You) *utility* scale. Let $\mathcal{L}$ be the (linear) space of all possible gambles on Ω, a *prevision* on Ω_X is defined as the expectation functional $\mathbb{E}_P(f) = \sum_{x\in\Omega_X} P(x)f(x)$, for some $f \in \mathcal{L}$.

Example 1. *Consider r.v. X and gamble f on event $\alpha \subseteq \Omega$. $f(\alpha) = \mathbb{I}_\alpha$, the indicator function, is such that a unit of money is won if $X \in \alpha$ occurs, 0 otherwise. For a given PMF P, $\mathbb{E}_P(\mathbb{I}_\alpha)$ is nothing but the price You are willing to place on gamble f. If You accept the transaction, $\mathbb{I}_\alpha - \mathbb{E}_P(\mathbb{I}_\alpha)$, Your expected gain is:*

$$(1 - \mathbb{E}_P(\mathbb{I}_\alpha))\, P(\alpha) + (0 - \mathbb{E}_P(\mathbb{I}_\alpha))\, (1 - P(\alpha)) = 0\,.$$

Let f be some gamble on Ω_X and let $\mathcal{D} \subseteq \mathcal{L}$ denote the set of *desirable gambles*. If f is bought at a price γ_*, gamble $f - \gamma_*$ results. This is desirable as long as the corresponding expectation is positive. We define $\underline{\mathbb{E}}_P(f)$ the supremum acceptable

[8] As well as insufficient data points and/or inconsistencies between different sources of evidence, and/or uncertainty about each source's reliability.

price, or rate, for buying gamble f [257], such that $\underline{\mathbb{E}}_P(f) = \sup_\gamma\{\gamma : f - \gamma \in \mathcal{D}\}$. Conversely, selling f for price γ^*, yields gamble $\gamma^* - f$, and $\overline{\mathbb{E}}_P(f)$ is the infimum acceptable price for selling gamble f; formally, $\overline{\mathbb{E}}_P(f) = \inf_\gamma\{\gamma : \gamma - f \in \mathcal{D}\}$. For fixed values $\underline{\mathbb{E}}_P(f), \overline{\mathbb{E}}_P(f)$, every gamble p is desirable as long as $p > \gamma_*$ or $p < \gamma^*$. If You have no preference between events, no bet at all is placed; and indeterminacy results.
Selling a gamble f for a price γ is equivalent to buying gamble $-f$ at price $-\gamma$:

$$\gamma - f = (-f) - (-\gamma) \tag{1.2}$$

Eq. (1.2) yields the lower-upper conjugacy relation:

$$\begin{aligned}\overline{\mathbb{E}}_P(f) &= \inf_\gamma\{\gamma : \gamma - f \in \mathcal{D}\} \\ &= \inf_\lambda\{-\lambda : -f - \lambda \in \mathcal{D}\} \\ &= -\sup\{\lambda : -f - \lambda \in \mathcal{D}\} \\ &= -\underline{\mathbb{E}}_P(-f)\,.\end{aligned}$$

Whenever lower and upper previsions coincide, i.e. $\underline{\mathbb{E}}_P = \overline{\mathbb{E}}_P = \mathbb{E}_P$, their corresponding value is called a *fair price* (under De Finetti's interpretation [257]), or a *linear prevision*. Linear previsions reproduce standard (finitely additive) probability models, that commit the gambler to buy f for any real price $p < \mathbb{E}_P(f)$, or to sell f at q, for any $q > \mathbb{E}_P(f)$. With lower previsions, f is bought at p for any $p < \underline{\mathbb{E}}_P(f)$, it is sold at q for any $q > \overline{\mathbb{E}}_P(f)$, and no decisions are taken for values in $\big(\underline{\mathbb{E}}_P(f), \overline{\mathbb{E}}_P(f)\big)$. Therefore, lower previsions allow *indecision* and may be intended as commitments to behave rationally. A behavior is called rational whenever it avoids sure loss, while being coherent. While avoidance of sure loss guarantees logical consistency - betting rates are accepted as long as they do not yield loss of utility under any circumstances -, logical closure, or coherence, implies full awareness of the implications that go along with betting rates.[9]
Formally, coherence of lower previsions requires:

1. $\underline{\mathbb{E}}_P(f) \geq \inf\{f(x) : x \in \Omega_X\}$, (Convexity)
2. $\underline{\mathbb{E}}_P(cf(X)) = c\underline{\mathbb{E}}_P(f(X)$, for every $c > 0$, (Positive Homogeneity)
3. $\underline{\mathbb{E}}_P\,(f(x) + g(x)) \geq \underline{\mathbb{E}}_P(f(X)) + \underline{\mathbb{E}}_P(g(X))$, (Super-Linearity)

where operations with gambles are intended as point-wise. If $\mathcal{D}$, set of desirable gambles, is coherent, it corresponds to a convex cone in $\mathcal{L}$, such that, if $f_1, \ldots, f_n \in \mathcal{D}$

[9] Coherence may be intended as *internal rationality* [183].

and $c_1, \ldots, c_n > 0$, $\sum_{i=1}^{n} c_i f_i \in \mathcal{D}$. In general, lower probabilities do not lead to coherent lower previsions [257, Sec. 2.7.3]. As for (sharp) probabilities, conditioning plays a substantial role in the reasoning process under uncertainty. We take up the so-called *updating interpretation* of conditioning, providing an understanding of Bayes rule as of a consistency requirement between current (unconditional) and conditional beliefs. See [272] on the temporal implications of conditioning in probabilistic reasoning, and their relationships with the theory of belief revision [3] (see Ch. 5).
Let $\mathcal{B}$ denote a partition of Ω, f is $\mathcal{B}$-measurable if it is constant on all the elements of $\mathcal{B}$. A *(separately) coherent conditional lower prevision* is defined as the functional $\underline{\mathbb{E}}_P(f|\mathcal{B})$ such that, for every $f, f' \in \mathcal{L}, \beta \in \mathcal{B}$ and $c > 0$, it holds:

1. $\underline{\mathbb{E}}_P(f|\beta) \geq \inf_{\omega\in\beta} f(\omega)$
2. $\underline{\mathbb{E}}_P(cf|\beta) = c\underline{\mathbb{E}}_P(f|\beta)$
3. $\underline{\mathbb{E}}_P(f + f'|\beta) \geq \underline{\mathbb{E}}_P(f|\beta) + \underline{\mathbb{E}}_P(f'|\beta)$

Consider now any target event $\alpha \subseteq \Omega$. If gambles are indicator of events, as in Ex. 1 (see p. 5), lower previsions are called lower (or upper) probabilities, denoted with $\underline{P}(\alpha)$ (or $\overline{P}(\alpha)$), measuring the strength of Your belief on event α. Again, the objectivist approach provides a behavioral interpretation to lower probabilities: as evidence supports event α, lower probabilities increase; symmetrically, as α is questioned by evidence, upper probabilities decrease. Avoidance of sure loss with lower probabilities implies $\underline{P}(\mathbb{I}_\alpha) \leq \overline{P}(\mathbb{I}_\alpha)$.

We may now introduce credal sets (CSs [157]) as flexible tools to represent uncertainty. Any CS is a convex set of PMFs, closed in the weak* topology[10]. Let X be any discrete random variable, $K(X)$ is defined by a set of PMFs P_X on Ω_X. A collection of PMFs over joint sets of variables, say $\mathbf{X}$, define a *joint* CS, $K(\mathbf{X})$. Any $K(X)$ may be equivalently specified by the collection of its *extreme points*, that we denote as $extK(X)$.[11]
An extreme point of a CS is any element of a closed and convex set that may not be derived as a convex combination of other PMFs. A CS is called *finitely generated* if $|extK(X)| < \infty$. Any such CS is defined by a set of linear constraints, and it thus may be equivalently represented as a polytope, i.e. the convex closure of a finite

[10]The weak* topology is the smallest topology such that all evaluation functionals $P(f)$ are continuous, where f is any gamble in the linear space $\mathcal{L}$.

[11]Whenever X is a binary random variable, its associated CS is fully specified by a single linear constraint of the type: $\underline{P}(x) \leq P(x) \leq \overline{P}(x)$, where $\Omega_X = \{x, \neg x\}$. This constraint is geometrically represented by a segment in the two-dimensional simplex space, with $|extK(X)| = 2$.

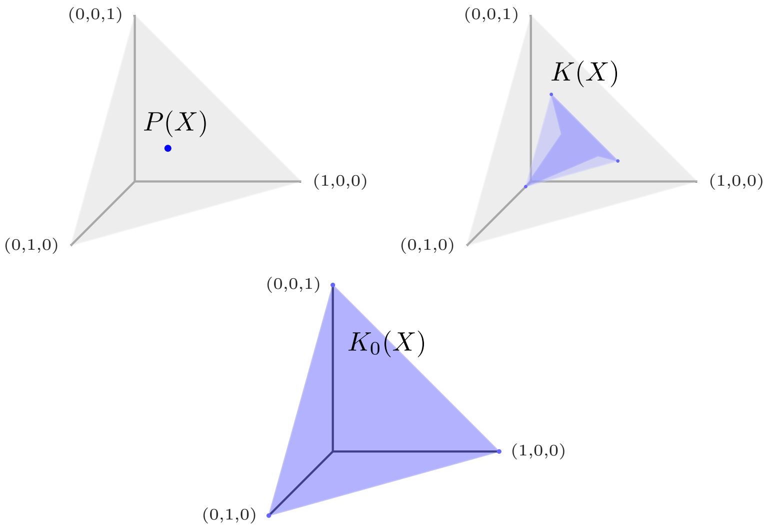


Figure 1.1: Geometric representation of a sharp (left panel) and imprecise (middle and right panels) probabilities in the three-dimensional probability simplex.

number of points, in the space of probability measures [51]. Let CH denote the convex-hull operator, it holds $K(X) = CH\{extK(X)\}$. Any two CSs are *equivalent* ($\equiv$) whenever their induced convex-hulls coincide, with no requirement on convexity. A CS that coincides with the probability simplex is called *vacuous*, and denoted with $K_0(X)$. Formally,

$$K_0(X) = \left\{ P(x) : P(x) \geq 0, x \in \Omega_x, \sum_{x \in \Omega_X} P(x) = 1 \right\} .$$

Example 2. *Fig. 1.1 provides a geometric representation of a single PMF (left panel), a CS (middle panel) and a vacuous CS (right panel) in the three-dimensional probability simplex. A sharp PMF corresponds to a single point in* Δ^3*, namely* $P(X) = (0.33, 0.33, 0.34)$*. The middle panel depicts a CS* $K(X)$*, defined by six extreme points;* $K(X)$ *is fully contained by the set (shaded blue area) induced by the lower and upper probabilities for each state of r.v.* X*. Its extreme points are:*

$$extK(X) = \left\{ \begin{array}{l} P_1(X) = (0.6, 0.2, 0.2), P_2(X) = (0.2, 0.6, 0.2), P_3(X) = (0.3, 0.4, 0.3), \\ P_4(X) = (0.2, 0.2, 0.6), P_5(X) = (0.5, 0.25, 0.25), P_6(X) = (0.6, 0.2, 0.2) \end{array} \right\} .$$

The vacuous CS $K_0(X)$ *trivially coincides with* Δ^3*.*

With CSs, indifference is specified by uniform probability models, whereas ignorance is implied by absence of linear constraints, i.e. by vacuous sets. Remarkably, CSs fully generalize belief functions.[12]

[12]We refer the interested reader to [7] for a full discussion on the subject.

Consider the specification of a CS, defined as a set of finitely additive probabilities, by a lower prevision:

$$K^{\underline{\mathbb{E}}_P}(X) = \{P \in \mathcal{P}_X : \mathbb{E}_P(f) \geq \underline{\mathbb{E}}_P(f), \forall f\} \ .$$

$\underline{\mathbb{E}}_P$ avoids sure loss if and only if $K^{\underline{\mathbb{E}}_P}(X)$ is not empty. Also, it is coherent if and only if $\underline{\mathbb{E}}_P$ is the lower envelope of $K^{\underline{\mathbb{E}}_P}(X)$, that is $\underline{\mathbb{E}}_P = \min_{P \in K^{\underline{\mathbb{E}}_P(X)}} P(X)$. Conversely, for a given credal set, lower previsions are derived as

$$\underline{\mathbb{E}}_P(f) = \min_{P(X) \in K(X)} \sum_{x \in \Omega_X} P(x) f(x)$$

Hence coherent lower previsions are in a one-to-one correspondence with CSs [54, 183, 48], and two equivalent CSs generate the same lower previsions. Extension to the conditional case is straightforward.
Consider the lower-upper conjugacy of indicator functions for event α, as above. By definition, $\underline{P}(\alpha) = \min_{P(X) \in K(X)} \sum_{x \in \alpha} P(x)$. Then,

$$\begin{aligned} 1 - \underline{P}(\alpha) &= \max_{P(X) \in K(X)} \left[1 - \sum_{x \in \alpha} P(x) \right] \\ &= \max_{P(X) \in K(X)} \sum_{x \notin \alpha} P(x) \\ &= \overline{P}(\neg \alpha) \ . \end{aligned}$$

When equality holds in the self-conjugacy relation, the CS is degenerate and it is defined as a linear functional.

Example 3. *Let X be any k-valued random variable, $k \geq 2$. It is defined on the k-dimensional probability simplex, denoted as Δ^k. If no linear constraints are provided, $K_0(X) = \Delta^k$ models ignorance. If a collection of constraints in form of lower-upper bounds $[p_{x,l}, p_{x,u}]$ is provided, for all $x \in \Omega_X$, a consistent credal set is defined as*

$$K(X) = \left\{ P(X) : p_{x,l} \leq P(x) \leq p_{x,u}, P(x) \geq 0, \sum_{x \in \Omega_X} P(x) = 1 \right\} \ .$$

Whenever $K(X) \neq \emptyset$, avoidance of sure loss is guaranteed by definition, as it holds:

$$\sum_{x \in \Omega_X} p_{x,l} \leq 1 \leq \sum_{x \in \Omega_X} p_{x,u} \ .$$

Also, by coherence, bounds are tight, i.e.

$$\begin{aligned} \sum_{x' \in \Omega_X \setminus \{x\}} p_{x',l} + p_{x,u} \leq 1 \\ \sum_{x' \in \Omega_X \setminus \{x\}} p_{x',u} + p_{x,l} \geq 1 \ . \end{aligned}$$

This is known as reachability *[257]. When $k = 2$, a CS for a binary variable has always at most two vertices in the probability simplex: $extK(X) = \{[p_l, 1-p_l], [p_u, 1-p_u]\}$*

1.2.1 Operations with Credal Sets

Marginalization on Ω_X of $K(X,Y)$, denoted as $K^{\downarrow X}(X,Y)$, for any pair X and Y, is obtained as [15, Ch. 9]:

$$K^{\downarrow X}(X,Y) = CH\left\{P(x) : P(x) = \sum_{y \in \Omega_Y} P(x,y), \begin{array}{l} x \in \Omega_X \\ P(X,Y) \in extK(X,Y) \end{array}\right\}$$

$K(X|Y = y)$ is the collection of all PMFs obtained by conditioning (cfr Eq. (1.1)) on each element of $K(X,Y)$, provided that $\underline{P}(y) > 0$, for every $P(Y) \in K^{\downarrow Y}(X,Y)$. Given a joint CS, *conditioning* is obtained by such *generalized Bayes rule* (GBR) [257, Sec. 6.4], and $K(X|Y) = \{K(X|y) : y \in \Omega_Y\}$[13]. $K(X)$ and $K(X|Y)$ are called *jointly coherent* if the second is the unique solution to GBR and if *conglomerability* holds.[14]

When $K(X|Y)$ and $K(Y)$ are given, we derive the joint CS by *composition*, or *combination*, with $\otimes$ denoting the associate operator:

$$K(X|Y) \otimes K(Y) = CH\left\{P(x,y) : P(x,y) = P(x|y)P(y), \begin{array}{l} x \in \Omega_X \\ P(X|y) \in extK(X|y) \\ y \in \Omega_Y \\ P(Y) \in extK(Y) \end{array}\right\}$$

Any sharp conditional PMF is represented by a conditional probability table (CPT), whose columns correspond to the conditioning values. A credal CPT (CCPT) may be either defined as *separately specified* (SS) CS, i.e. $K(X|Y) = \{K(X|y) : y \in$

[13]With coherent lower previsions, $\underline{\mathbb{E}}_P\left(f(X)|y,z\right)$ is defined as the unique solution to the GBR, defined as:

$$\underline{\mathbb{E}}_P\left[\left(f(X) - \lambda\right)\mathbb{I}_y(Y)|z\right] = 0$$

whenever $\underline{P}(y|z) > 0$, otherwise no unique solution may be found; see [259, 55] for discussion and proposals to this latter setting.

[14]When Ω_Y is a finite partition of $\Omega_{\mathbf{X}}$, conglomerability follows from coherence of $K(X|Y)$ and GBR. Formally, for a given partition $\mathcal{B}$ of Ω, a coherent lower prevision is $\mathcal{B}$-conglomerable if there exists a $\underline{\mathbb{E}}_P(.|\mathcal{B})$ such that $\underline{\mathbb{E}}_P(.)$ and $\underline{\mathbb{E}}_P(.|\mathcal{B})$ are coherent. When unconditional and conditional previsions are linear, *B-disintegrability* applies. A disintegrable linear prevision is always conglomerable, whereas the converse is not true in general [91]. Joint coherence, from a behavioral perspective, implies that a finite combination of desirable gambles, according to $\underline{\mathbb{E}}_P(.)$ and $\underline{\mathbb{E}}_P(.|\mathcal{B})$, is itself desirable [257, Ch. 6].

$\Omega_Y\}$, or *nonseparately specified* (nSS). The *extensive* specification of a nSS conditional credal set yields an extensive CPT (ECPT), defined as a finite collection of CPTs. Any CCPT may be transformed into an ECPT, by considering all possible combinations from the elements of the sets [171]. Conversely, an ECPT may be transformed into a CCPT; we will return to this in Sec. 1.3.2.

1.2.2 Independence

Given a joint CS $K(X,Y)$, consider $K(X|Y)$, obtained by GBR. For a given random variable Z, $K(X|y,z) \equiv K(X|z)$ if and only if $CH\{K(X|y,z)\} = CH\{K(X|z)\}$, for all $x \in \Omega_X, y \in \Omega_Y, z \in \Omega_Z$. As a first remark, symmetry is no longer an intrinsic property of independence with imprecise probabilities. Also, the concept of *irrelevance*, interchangeably used with independence in the standard setting [66], assumes its own prominence among sets of PMFs.

We hereby list a (not exhaustive) selection of concepts for conditional independence that are relevant to our purposes.[15] Analogously to Def. 3, we write $X * Y|Z$ to denote a $*$-irrelevance relationship.

Definition 4 (Epistemic Irrelevance (ER) [257]). *$(X\ ER\, Y|Z)$ if and only if it holds:*

$$K(x|z) \equiv K(x|y,z)\,,$$

for each $(x,y,z) \in \Omega_X \times \Omega_Y \times \Omega_Z$.

ER ought to be intended as Galles and Pearl's *informational irrelevance* [108].

Definition 5 (Epistemic Independence (EI) [257]). *$(X\ EI\, Y|Z)$ holds when X is irrelevant to Y given Z and vice versa, i.e. it is the symmetric version of ER [54]:*

$$(X\ EI\, Y|Z) \iff (X\ ER\, Y|Z) \wedge (Y\ ER\ X|Z)\,.$$

Given a joint set of PMFs, EI does not imply the first factorizes [53], although other desirable features are satisfied, such as marginalization, associativity and external additivity (see [77] for details).

Definition 6 (Strong Independence [16]). *Strong independence extends SI to imprecise probabilities. Straightforward extension of SI to sets of PMFs goes under the*

[15]See [258, 73, 49, 56] for details.

[16]According to [53], the term was introduced by Walley [256], yet with a slightly different approach to zero-lower probability events. Its introduction dates back to Levi [157], where it was referred to as *strong confirmational irrelevance*.

name of complete independence *which, by definition, fails convexity [53]. Under strong independence, SI applies to the extreme points of a credal set [186]: X and Y are strongly independent conditional on Z if and only if the following holds:*

$$extK(X|Y,z) = extK(X|z) \tag{1.3}$$

for every $z \in \Omega_Z$. This is equivalent to taking the convex hull of the set of PMFs obtained by complete independence [53].

Given any two random variables X and Y, their *strong extension* corresponds to $K^{SE}(X,Y)$, the CS obtained by strong independence (see Sec.1.3.2).
Different versions and terms have been used in [186] for strong independence, that we list below. However, when not explicitly stated, we assume variables X and Y are (unconditionally) strongly independent whenever Eq. (1.3) is satisfied.

1. Strong Independence on Distribution (St_1):

$$K(X,Y,Z) = \left\{ P(x,y,z) = \frac{p'(x,z)p''(y,z)}{p'(x,z)^{\downarrow Z}}, \begin{array}{c} p' \in K(x,z), \\ p'' \in K(y,z), \\ p'^{\downarrow Z}(x,z) = p'^{\downarrow Z}(y,z), \\ \forall(x,z), \forall(x,z) \end{array} \right\}.$$

   The induced decomposition operator fails the Shenoy-Shafer associativity axiom [226]. This issue is relevant to inference with CNs.

2. Strong Independence on Decomposition (St_2): St_2 requires $K(X,Y,Z) = K(X,Z) \otimes K(Y,Z)$, with $K(i,Z) \neq K(i,j,Z)^{\downarrow i,Z}$ in general[17]. The joint CS derived under St_1 contains that induced by StI_2 of X and Y conditional on Z [186, Th.4].

3. Causal Irrelevance (CE):

$$K(X,Y,Z) = K(X,Z) \otimes K(Y|Z)$$

   $K(X,Y,Z)$ is obtained by taking the product $P(x,z)P(y|z)$, for all $x \in \Omega_X, y \in \Omega_Y$ and $z \in \Omega_Z$, with $P(X,Z) \in K(X,Z)$ and $P(Y|Z) \in K(Y|Z)$. Element-wise CE (ECE) requires instead:

$$K(X,Y,Z) = \{K(X,Y,z) = K(X,Z) \otimes K'(Y|z) : z \in \Omega_Z\}$$

With standard probabilities, concepts of independence and irrelevance are equivalent. Independence, intended as *factorization*, yields modularity of a statistical

[17] Remember from Sec. 1.2.1 symbol $\otimes$ denotes composition.

	ER	EI	StI_1	StI_2	(E)CE
Symmetry		✓	✓	✓	
Decomposition	$\rightarrow, \leftarrow^*$	✓*	✓	✓	$\rightarrow, \leftarrow$
Weak Union	$\rightarrow^*, \leftarrow^*$	✓*	✓	✓	$\leftarrow$
Contraction	$\leftarrow$		✓		$\rightarrow$
Redundancy	$\rightarrow, \leftarrow$	✓	✓	✓	$\leftarrow$

Table 1.1: Semi-graphoid properties satisfied by different concepts of independence for imprecise probabilities. The symbol ✓ denotes fulfillment of a graphoid axiom, $\rightarrow$ ($\leftarrow$) is used to denote fulfillment of direct (reverse) conditions, superscript $*$ restricts fulfillment of the property to positivity of conditioning events, defined as strictly positive whenever (y, z) is possible, for every $y \in \Omega_Y, z \in \Omega_Z$ [54]. Proofs may be found in [54, Th. 1, 3] (for ER and EI, respectively), [186, Th. 3, 5, 6, 7] (for StI_1, StI_2, CE and ECE, respectively).

model, that in turn allows reduced parametric dimensionality and, as a consequence, computational tractability [194]. Also, Markov condition and graphoid axioms provide means for a causal approach to independence [234]. With imprecise probabilities SI implies EI; and EI, in turn, trivially implies ER. Since irrelevance relationships are now asymmetric, direct and reverse ($\rightarrow$ and $\leftarrow$, respectively) conditions for graphoids must be considered. Additional forms other than direct and reverse exist[18], yet they are not fulfilled by most concepts of independence, including those hereby considered. Table 1.1 summarizes the (semi-)graphoid properties satisfied by the independence concepts introduced above.

Finally, a concept worth mentioning, related to independence with imprecise probabilities, is that of *dilation* [260]. This occurs when conditioning on some event renders knowledge of variable X *more imprecise*, formally: $K(X) \subset K(X|y)$, for all events $y \in \Omega_Y$, this latter inducing a partition of Ω. See [196] for a thorough description of dilation and details on its (strong) relationship with independence.

1.3 Probabilistic DAG-Models

Probabilistic Graphical Models (PGMs) compactly represent the multivariate joint behavior of a collection of random variables, based on their (possibly) complex pattern of independence relationships [68]. Any PGM is specified by i) a *graph*, and

[18]Particularly, graphoid axioms are considered by interchanging all elements in their definitions from Def. 3. This way, redundancy has two versions (right and left [108], or direct and reverse [54]), decomposition and weak union have four, contraction and intersection have eight.

ii) a *probabilistic* component. Additionally, if the latter is not sharp, specification of iii) a *concept of irrelevance* is required for a PGM. We hereby outline fundamental concepts from graph theory, before we formally introduce Bayesian and Credal Networks as PGMs based on a shared class of graphs.

Basic Concepts of Graph Theory

A graph is a mathematical structure, specified by the ordered pair $(\mathbf{V}, \mathbf{E})$. The first element of the pair, $\mathbf{V} = \{X_0, \ldots, X_n\}$, is a collection of objects, called *nodes*, with $n \geq 0$, while $\mathbf{E}$ is and a collection of arcs and/or edges, linking pairs of nodes. Nodes are graphically represented as circles, arcs as segments $(-)$ and edges as oriented arrows $(\rightarrow)$.

A graph is called *undirected* if $\mathbf{E}$ only contains arcs. An arc, linking node X_i to X_j, is denoted with $((i, j))$. Let $Adj(X_i)$ be the set of nodes adjacent to X_i, $((i, j)) \in \mathbf{E}$ implies $X_i \in Adj(X_j)$ and $X_j \in Adj(X_i)$.

A maximally connected set of k nodes is called a *clique* of size k, $k \geq 2$. Elements of $\mathbf{X}$ form a clique whenever they are pairwise adjacent: $((i, j)) \in \mathbf{E}$, for each pair $(X_i, X_j) \subseteq \mathbf{X}$. Trivially, a pair of adjacent nodes constitutes a clique of size 2.

In a directed graph, edges are denoted as (i, j), such that $(i, j) \in \mathbf{E}$ implies X_i is a *parent* of X_j, and simmetrically, X_j is a *child* of X_i. Formally, $X_i \in Pa(X_j)$ (*parent set* of X_i) and $X_j \in Ch(X_i)$ (*set of children* of X_i), respectively. A node with no parents is called a *root*; it is called a *leaf* if it has no children.

Any two nodes X_i and X_j in a directed graph are connected by a (directed) *path* π_{ij} if there exists a sequence of nodes $< X_i, \ldots, X_j >$, whose directed edges may be walked from the first element to the last. An acyclic directed graph (DAG) is a directed graph such that, if two nodes are connected by π_{ij}, then no π_{ji} exists. As a trivial consequence, $(i, j) \in \mathbf{E}$ implies $(j, i) \notin \mathbf{E}$. We denote any DAG as $\mathcal{G}$.

Let $\mathbf{X} \subseteq \mathbf{V}$, its induced *subgraph* $\mathcal{G}^{\mathbf{X}}$ has set of nodes $\mathbf{X}$ and edges $\mathbf{E}_{\mathbf{X}} = \mathbf{E} \cap (\mathbf{X} \times \mathbf{X})$.

Whenever two non-adjacent nodes share a neighbor, i.e. an adjacent node, their ensemble is called an *unshielded triple*. In a DAG, an unshielded triple $< X, Y, Z >$ such that X and Y have child Z in common is called a *v-structure*, and Z is a *collider*; see Fig. 1.2 (left panel). A DAG is called *singly connected*, or a *polytree*, if there exists at most one path connecting each pair of nodes in $\mathbf{V}$, otherwise it is *multiply connected*. d is used to denote the *treewidth* of a DAG, defined as its maximum indegree, this latter defined as the number of incoming edges:

$$d = \max_{i=0,\ldots,n} |Pa(X_i)|$$

A singly connected DAG is a *tree* whenever $d = 1$.

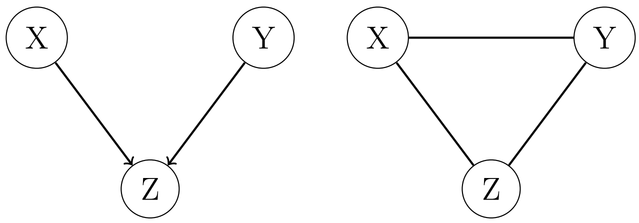


Figure 1.2: Left panel: v-structure; Right panel: moralized v-structure.

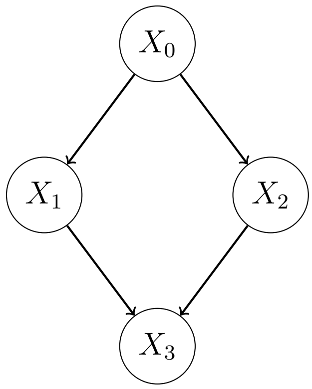


Figure 1.3: A toy multiply connected Bayesian network, with $\mathbf{V} = \{X_0, X_1, X_2, X_3\}$.

Example 4. *Consider Fig. 1.3.* $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, *whose components correspond to:*

$$\begin{aligned} \mathbf{V} &= \{X_0, X_1, X_2, X_3\}, \\ \mathbf{E} &= \{(X_0, X_1), (X_0, X_2), (X_1, X_3), (X_2, X_3)\}. \end{aligned}$$

X_0 *is a root node, i.e. it has empty parent set, whereas* X_3 *is a leaf. Also,* X_3 *is a collider, since* $Pa(X_3) = \{X_1, X_2\}$ *and there does not exist edge* (X_1, X_2), *nor* (X_2, X_1), *in* $\mathbf{E}$; $< X_1, X_3, X_2 >$ *is a v-structure.* $\mathcal{G}$ *is multiply connected, with treewidth* $d = 2$.

We shall now introduce DAG-based models. We refer the reader interested in details on graph theory to [190, Ch.6-8] for a complete introduction.

1.3.1 Bayesian Networks

Bayesian Networks (BNs [194]) are probabilistic graphical models defined over a set of r.v.s $\mathbf{V} = \{X_0, \ldots, X_n\}$. A BN $\mathcal{B}$ is specified by the pair $(\mathcal{G}, P)$.
P corresponds to the probabilistic component of the PGM. At this stage, we consider only PMFs that are strictly positive over the sample space induced by the joint r.v. $\mathbf{V}$, although we will drop such a requirement later in the present work. Elements of $\mathbf{V}$ are in one-to-one correspondence with the nodes of the graphical

component of $\mathcal{B}$. This is specified as a DAG $\mathcal{G}$ over $\mathbf{V}$. As a remark, since every random variable X_i is in one-to-one correspondence with a node in the DAG, same notation and terminology will be used throughout interchangeably, when referring to either $\mathcal{G}$ or P. See [150] for an overview of PGMs based on different graphical components.
A BN is called *discrete* [127] if each $X_i \in \mathbf{V}$, $i = 0, \ldots, n$, and $\mathbf{V}$ are Multinomial r.v.s. Other types of DAG-based PGMs were proposed in the literature; among others, *Gaussian* BNs [127], where $\mathbf{V}$ is a multivariate Normal r.v., resulting from the linear combination of (univariate) Normal r.v.s, *Conditional Linear Gaussian* [155] and *Copula* [96] BNs. We consider Discrete BNs only throughout.
By the so-called *Markov Condition*, each node X_i is independent of its non-descendants in $\mathcal{G}$, conditional on its parents. For a chosen concept of irrelevance[19] for P, this implies graphical separation, e.g. d-separation. This was introduced by Pearl in [194] as a formal procedure for detecting conditional SI relationships, implied by the Markov condition:

Definition 7 (D-separation [194]). *Any two nodes X and Y are d-separated by $\mathbf{Z}$ in $\mathcal{G}$ if every undirected path connecting the two, i.e. any sequence of edges whatever their direction, is* blocked *by $\mathbf{Z}$. Without loss of generality, let $\mathbf{Z} = \{Z\}$ and consider unshielded triple $< X, Z, Y >$. The path connecting X to Y is blocked if either:*

1. *The triple is either a Markov chain ($X \to Z \to Y$) or a fork path ($X \leftarrow Z \to Y$), and Z is instantiated to some value, i.e. it is observed.*

2. *The triple is a v-structure and Z is not observed.*

In the general case, consider undirected path π_{ij}^u (π_{ij} with dropped directions) and $|\mathbf{Z}| \geq 1$; π_{ij}^u is blocked if all nodes from $\mathbf{Z}$ that are instantiated are not colliders in π_{ij}.

Remarkably, d-separation satisfies the graphoid axioms [153].[20]

Any PMF P over $\mathbf{V}$ is *faithful* to $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ if all conditional independence relationships among pairs of variables in the PMF are encoded by the graph; see, e.g. [234, 150], for further details. When P is faithful to $\mathcal{G}$, it satisfies the *Faithfulness condition*.
If both Markov and Faithfulness conditions are satisfied, P may be equivalently specified as the collection of $(n+1)$ conditional probability tables (CPTs) $\{P(X_i|Pa(X_i))$:

[19] Remember concepts of irrelevance may all be reduced to SI, with strictly positive sharp PMFs.
[20] See [247] on a-graphoid L-separation for a class of graphical models based on cs-independence [248], accounting for zero-probability conditioning events.

$i = 1, \ldots, n\}$, called *local* distributions. The CPT of random variable X_i has $|\Omega_{X_i}|$ rows and $\max(1, |Pa(X_i)|)$ columns, $i = 1, \ldots, n$.
It holds [194]:

$$P(X_0 = x_0, \ldots, X_n = x_n) = \prod_{i=0}^{n} P(X_i = x_i | Pa(X_i) = pa(X_i)), \tag{1.4}$$

where $pa(X_i) \in \Omega_{Pa(X_i)}$ is consistent with configuration $(x_0, \ldots, x_n) \in \Omega_{\mathbf{V}}$.

Example 5 (Ex. 4 continued)**.** *$\mathcal{B}$ is a BN whose joint PMF may be represented as:*

$$P(x_0, x_1, x_2, x_3) = P(x_0)P(x_1|x_0)P(x_2|x_0)P(x_3|x_1, x_2)$$

for every $(x_0, x_1, x_2, x_3) \in \Omega_{\mathbf{V}}$, provided $P(x_0), P(x_1, x_2)$ are strictly positive.

Indeed, parametrization of Gaussian models requires specification of partial correlation coefficients between each node and its set of parents, rather than the elements of CPTs.
Let us stress the following: for any v-structure $< X, Z, Y >$ it holds[21]

$$I(X, Y), \quad \neg I(X, Y; Z).$$

Whenever a node E is instantiated, i.e. it is observed, $P(\mathbf{V})$ is changed to $P(\mathbf{V} \setminus \{E\}|e)$.
An equivalent representation of the updated BN requires columns of the CPTs of E's children that are consistent with observation e are selected, and outgoing edges of E may be removed.
A weaker concept of graphical independence was introduced by Boutilier for BNs [26], based on the following:

Definition 8. *Random variables X and Y are CSI-separated by $\mathbf{Z} = \mathbf{z}$ if they are d-separated in the equivalent DAG induced by observation $(\mathbf{Z} = \mathbf{z})$.*

Definition 9 (Context-Specific Independence in Bayesian networks)**.** *Random variables X and Y are CSI (see Def. 2) conditional on $\mathbf{Z} = \mathbf{z}$ if and only if they are CSI-separated by $\mathbf{Z} = \mathbf{z}$.*

As an example, suppose a BN is used to describe the average income of a worker, given its job and the monthly weather conditions. Trivially, knowledge of the latter is expected to have a null (or spurious at most) influence on the income of an office worker, whereas it will be significantly associated with a farmer's. CSI patterns

[21] Remember term $I(X, Y; Z)$ was introduced in Def. 1 to denote SI of X and Y conditional on Z.

allow further reduction of the parametric dimensionality of a given BN.
CSI representation in dependency models was also considered by [112, 26, 199]. Particularly, Pensar *et al.* [199] introduced *labeled* DAGs (LDAGs), that compactly represent *local* CSIs, i.e. CSI among a subset of parents and their child, conditional on the remaining parents. CSI-separation was proved to be a sound, yet not complete, method for detecting such independence relationships in LDAGs [199]. In Sec. 2 of the Appendix, parameter dimensionality reduction CSI will be considered within the framework of structural learning, that we shall now briefly outline.

The following sections provide an overview of the typical tasks with BNs. With PGMs in general, these are characterized as either learning or inference tasks. The former include model identification, i.e. choice of the graphical component, and/or estimation of its parameters. In a Discrete BN, this requires compilation of each column of each node's CPT. Learning tasks are called *structural* and *parameter*, respectively, when they tackle estimation of $\mathcal{G}$ and P's parametrization. Probabilistic reasoning with a given PGM is usually referred to as *inference*, given a fully specified model.[22]

Bayesian Networks: Learning

Given a dataset, *structural learning* (SL) requires estimation of the graphical component of a BN, i.e. of $\mathcal{G}$. SL may occur following either a *constraint-based* or a *search-and-score* approach, that we briefly characterize.
Algorithms for constraint-based structural learning of a BN make use of conditional independence tests, to assess presence (or absence) of an edge in the graph.[23] Among others, the PC algorithm [234] iteratively removes arcs from a complete undirected graph, where each node is adjacent to all others, by performing independence tests over adjacency sets of increasing cardinality. Unshielded triples are eventually oriented as v-structures, and a partially directed acyclic graph results. Any partially directed graph induces a *Markov equivalence class* [234], defined as a collection of conditional independence statements[24]. Finally, arcs are oriented, based on Zhang's rules [273]. The oracle PC algorithm is a sound and complete procedure if and only if both Markov and Faithfulness conditions hold, whereas its sample version also requires the chosen independence test to be correct.

[22] This terminology is classical in the theory of PGMs, yet we reckon it is rather counterintuitive in the statistical literature.

[23] See, e.g. [249, 234, 146].

[24] Markov chains and fork paths induce the same independence statements, and orientation of the edges may be neglected: $X \to Y \to Z$, $X \leftarrow Y \leftarrow Z$ and $X \leftarrow Y \to Z$ belong to the same Markov equivalence class, where it holds: $I(X, Z; Y)$.

In [209], Ramsey extended the algorithm above to the case of weaker forms of Faithfulness, by allowing the routine to mark ambiguous unshielded triples as *unfaithful*. The graph resulting from Ramsey's conservative PC (CPC) algorithm is called an *e-pattern*; this is no longer a single Markov equivalence class as its associated independence statements are consistent with several classes. The CPC algorithm will be considered in Sec. 2 of the Appendix.
Search-and-score algorithms, on the other hand, explore the space of all possible DAGs over $\mathbf{V}$. They tackle optimization of a scoring function, reporting the (penalized) goodness of fit of a graph, given data. Search is usually based on heuristics, to make the procedure efficient. See, e.g. [127, 105, 189]. Among others, *hill-climbing* ([217]) improves *greedy* search of the space of graphs by local perturbations on candidate structures. Also, Friedman's Structural EM algorithm [105] is worth mentioning, tackling the case of incomplete records of data.
All structural learning processes may be partially constrained upon knowledge-based domain. This possibly involves *whitelisting* [220] edges, making them always result in candidate solutions. *Blacklisted* learning procedures work conversely.
If the pattern of conditional independence relationships is already available, *parameter learning* of a BN may occur following either a maximum likelihood or a Bayesian approach [105, 39, 220]. Also, qualitative as well as quantitative constraints, e.g. across the columns of a CPT, may be posed while estimating the parameters of a BN from data; see [191]. When data are missing, i.e. records are incomplete, the EM algorithm may be considered for parameter estimation. The EDML algorithm was introduced as an alternative (efficient) approach by Choi *et al.* in [39]. Their proposal is based on the optimization of a collection of likelihood functions, resulting from the propagation of probabilistic evidence, properly specified, through auxiliary graphical structures, called *meta-networks* [64]. We will return to this approach in Ch. 3, where propagation of probabilistic evidence in BNs (and their credal counterpart) is considered. Finally, a robust learning approach with missing data was proposed by Ramoni and Sebastiani [207]; see also [83].

Bayesian Networks: Inference

We hereby introduce next basic inferential tasks, or queries, that may be considered with BNs.[25]
Let α be any *target* event in $\Sigma = 2^{\Omega}$,

1. Simple queries:

[25] See [150] for a detailed list.

a) $\alpha = (x_0, \ldots, x_n)$, for any $(x_0, \ldots, x_n) \in \Omega$. $P(\alpha)$ is computed as from Eq. (1.4).

b) Let $\mathbf{X}_Q \subseteq \mathbf{V}$, $\alpha = \mathbf{x}_Q$, for any $\mathbf{x}_Q \in \Omega_{\mathbf{X}_Q}$; computation of $P(\alpha)$ requires *marginalization* of $P(x_0, \ldots, x_n)$ over $\mathbf{V} \backslash \mathbf{X}_Q$, whose generic element is $\mathbf{x}_{-Q}$:

$$P(\mathbf{x}_Q) = \sum_{\mathbf{x}_{-Q} \in \Omega_{\mathbf{V} \backslash \mathbf{X}_Q}} \prod_{i=0}^{n} P(x_i | pa(X_i)) \tag{1.5}$$

provided $(x_0, \ldots, x_n)$ is *consistent* ($\sim$) with target event $\mathbf{x}_Q$.

c) Let α be any conditional event $\mathbf{x}_Q|\mathbf{e}$, $\mathbf{e} \in \Omega_{\mathbf{E}}$, $\mathbf{E} \subseteq \mathbf{V}$. $P(\alpha)$ is obtained by *conditioning*, or *updating*, on observed evidence $\mathbf{e}$:

$$P(\mathbf{x}_Q|\mathbf{e}) = \frac{\sum_{\mathbf{x}_{-(Q,\mathbf{E})} \sim (\mathbf{x}_Q, \mathbf{e})} \prod_{i=0}^{n} P(x_i|pa(X_i))}{\sum_{\mathbf{x}_{-\mathbf{E}} \sim \mathbf{e}} \prod_{i=0}^{n} P(x_i|pa(X_i))} \,. \tag{1.6}$$

Eq. (1.6) is just Bayes rule, and the term at the denominator is the *probability of evidence*[26]. Exact BN updating is NP-hard with general DAGs (although polynomial-time schemes are available for polytrees[27]).

2. Decision tasks:

d) Let r be any rational number, and $\mathbf{X}_Q$ be a collection of target variables and $\mathbf{E} = \mathbf{e}$. For any fixed $\mathbf{x}_Q \in \Omega_{\mathbf{X}_Q}$, is it true:

$$P(\mathbf{x}_Q|\mathbf{e}) > r\,? \tag{1.7}$$

e) Maximum A Posteriori (MAP): Find $\mathbf{x}_Q \in \Omega_{\mathbf{X}_Q}$ such that[28]:

$$\mathbf{x}_Q^* = \mathrm{argmax}_{\mathbf{x}_Q \in \Omega_{\mathbf{X}_Q}} P(\mathbf{x}_Q|\mathbf{e}) \,. \tag{1.8}$$

MAP is NP-complete on singly connected networks, it is NP^{PP}-complete with general DAGs. When $\mathbf{X}_Q = \mathbf{V} \backslash \mathbf{E}$, MAP is called Most Probable Explanation (MPE). MPE requires finding the complete instantiation $(x_0, \ldots, x_n)$ consistent with $\mathbf{e}$, such that $\underline{P}(x_0, \ldots, x_n) > r$, for some $r > 0$. Such task is polynomial with singly connected networks, otherwise it is NP-complete.

[26] So far, we assumed P is strictly positive and no requirements are thus necessary. In the general case, standard conditioning requires $P(\mathbf{e}) > 0$

[27] See, e.g. [150] for a review.

[28] Additional tasks exist in the literature, such as the *maxmin* problem, for a fixed configuration of a subset of elements of $\mathbf{X}_Q$; see [144].

3. Sensitivity analysis: Let α be some event, $\mathbf{x}_Q$ be the target event, and r and $\mathbf{E} = \mathbf{e}$ as before. Find the minimum change $\delta_{\mathbf{x}_Q|\mathbf{e}}$ such that:

$$P(\alpha|\mathbf{x}_Q, \mathbf{e})P(\mathbf{e}) \left[P(\mathbf{x}_Q|\mathbf{e}) + \delta_{\mathbf{x}_Q|\mathbf{e}}\right] > r$$

We introduce the following:

Definition 10 (m-irrelevance)**.** *For any $\mathcal{B}$, X is m-irrelevant to Y whenever Y's PMF is invariant to changes in the behavior of X:*

$$\sum_{x\in\Omega_X} P(y|x)P(x) = \sum_{x\in\Omega_X} P(x|y)\left[P(x) + \delta_x\right]$$

for any $\delta_x \in \mathbb{R}$, as far as $\sum_x [P(x) + \delta_x] = 1$.

Unobserved leaves are always m-irrelevant to the remaining nodes in $\mathcal{G}$. M-irrelevance may be readily detected by, e.g. the Bayes Ball algorithm [223].
Exact approaches are based on enumeration sequences, aimed to efficiently solve a combinatorial optimization problem. Among others, Lauritzen and Spiegelhalter proposed building a *Junction Tree* (JT) from the DAG, whose nodes are cliques [156]. Other approaches include exact message-passing procedures [194], differential approaches based on reformulation of the joint PMF as a multilinear function [63], recursive decomposition [46], optimal factoring [161], symbolic probabilistic [62] and causal independence based [275] inference methods. In the following, we provide a sketch of the Variable Elimination (VE) method [81], to provide an intuition.
Consider marginalization (inferential task b). Inference methods differ in the approach they follow when summing out variables in $\mathbf{V}\backslash\mathbf{X}_Q$. With VE, an ordering for the variables is chosen, with target variables as last. For every r.v. X in the ordering, all local PMFs that contain it are pooled into a *bucket* of X. All CPTs in X's bucket are removed from the list of all CPTs for the network, and the product of all CPTs in the bucket marginalizes X out. The resulting PMFs are inserted in the list of all CPTs for the network, and the next random variable from the ordering is selected.

Example 6. *Consider Fig. 1.3 and let $\{X_0, X_1, X_2, X_3\}$ be the fixed ordering, with target node $\mathbf{X}_Q = \{X_3\}$. Following VE, the list of all CPTs for the BN contains $P(X_0) = \{P(x_0) : x_0 \in \Omega_{X_0}, \sum_{x_0\in\Omega_{X_0}} P(x_0) = 1\}$, $P(X_1|X_0) = \{P(x_1|x_0) : x_1 \in \Omega_{X_1}, \sum_{x_1\in\Omega_{X_1}} P(x_1|x_0) = 1, x_0 \in \Omega_{X_0}\}$; $P(X_2|X_0)$ and $P(X_3|X_1, X_2)$ defined analogously.*
Root node X_0 is selected first, its bucket containing $P(X_0), P(X_1|X_0), P(X_2|X_0)$.
r.v. X_0 is marginalized out: $\sum_{x_0\in\Omega_{X_0}} P(x_0)P(X_1|x_0)P(X_2|x_0) = P(X_1, X_2)$; the updated list reads: $P(X_1, X_2)$, $P(X_3|X_1, X_2)$.

Selection of r.v. X_1 *yields* $\sum_{x_1 \in \Omega_{X_1}} P(x_1, X_2)P(X_3|x_1, X_2) = P(X_3, X_2)$.
Successively, r.v. X_2 *is selected:* $\sum_{x_2 \in \Omega_{X_2}} P(X_3, x_2) = P(X_3)$.
The list now contains $P(X_3)$ *only, and* $P(x_3)$ *may be computed for any* $x_3 \in \Omega_{X_3}$.

Exact inference is always NP-hard on general networks [45] and it becomes infeasible, in space and/or time, as the cardinality of both $\mathbf{V}$ and $\mathbf{E}$ increases. Approximate techniques are also NP-hard [61]. We distinguish them into three main classes:

Model Simplification The original problem is reduced toward a simplified graphical and/or probabilistic structure, and exact inference is performed. Among others, they include (loopy) Cutset conditioning. See, among others [261, 143, 274] and, e.g. [64] for a review.

Search-and-Score Analogously to the learning setting, these methods search for the most likely partial instantiations. See, e.g. [130, 275, 218].

Simulation Algorithms Stochastic simulation is used to randomly select configurations from the sample space of $\mathbf{V}$, based on either Monte Carlo methods or importance sampling. Frequencies are computed as an approximation; see [5], and references therein, for a survey. Among others, systematic Latin Hypercube sampling was considered by [38]. An application of Latin Hypercube sampling for the estimation of an over-parametrized complex compartmental model may be found in Sec. 1 of the Appendix.

1.3.2 Credal Networks

Credal networks (CNs) [51] extend BNs to the imprecise framework, and provide a compact representation for sets of PMFs. They assign a local CS to each node in a DAG. Like BNs, they are used to support decision making.
A single credal network represents all sharp PMFs (or equivalently BNs) that are consistent with the set of linear constraints used to specify the first. A CN may be intended as a BN whose parameters (one, some or all) are not known *precisely*, but rather they are defined by convex constraints. A CN with a single credal node is called *near-Bayesian*. CNs are not the only graphical model that has been proposed in the literature for modeling and reasoning under imprecision; e.g. [233], and [184] as a generalized representation for CNs (and hence for BNs).

Formally, a CN is defined by the triple $(\mathcal{G}, \{K(X|Pa(X)) : X \in \mathbf{V}\}, \mathcal{I})$. The second element is a collection of locally specified CSs, and $\mathcal{I}$ is the adopted independence concept. Unlike BNs, there is not unique way to combine locally defined CSs

and, from previous section, independence concepts are no longer equivalent with imprecise probabilities. Several *extensions* may thus be considered for a CN, based i) on the adopted independence concept, and ii) on the chosen *combination* rule.
Provided a fixed irrelevance concept, say $*$, the Markov Condition for CNs requires X to be $*$-irrelevant to its non-descendants non-parents, given its parents. The largest joint CS on $\mathbf{V}$, consistent with the collection of local CSs, corresponds to the *Natural Extension* of the network, denoted as $K^{NE}(\mathbf{V})$.
Suppose Fig. 1.3 is a credal network. The linear constraints of a credal network are locally specified and refer to single nodes. If a CS is separately specified, those constraints involve single conditional PMFs, otherwise they are not-SS. That is, if all constraints are of the form $P(x_2|x_0) \geq k \cdot P(\neg x_1|x_0)$, they induce a separately specified credal network, whereas presence of constraints of the type $P(x_1|x_0) \geq k \cdot P(x_1|\neg x_0)$ yield non-separately specified ones. Non-separate CNs have received little attention in the literature, compared to separate ones, with the notable exception of the credal versions of the Naïve Bayes Classifier (NBC) and Tree Augmented NBC (TAN) [47]. However, any non-separate CN may be equivalently specified as a separate CN on a larger domain, by properly introducing auxiliary nodes that *separate* non-separate CPTs [11].
Under strong independence, all extreme distributions of a CN factorize as a BN. Let $K^{SE}(\mathbf{V})$ denote the strong extension (SE) of a CN, whose extreme points are $extK^{SE}(\mathbf{V}) = \{P_j(\mathbf{V}) : j = 1, \ldots, \nu\}$[29]. It holds:

$$P_j(x_0, \ldots, x_n) = \otimes \prod_{i=0}^{n} p(x_i|pa(X_i)) \,, \tag{1.9}$$

where $\otimes$ is the composition operator, already defined, and each p is an extreme point of its associate CS K, for any $(x_0, \ldots, x_n) \in \Omega$.
Strong Extension is derived based on the Markov Property for different concepts of Strong Independence [186], namely StI_1, StI_1, CE, ECE. For a given CN with all variables logically independent, (graphical) d-separation implies strong independence and EI in Strong Extensions [52].

Credal Networks: Inference

Inference in credal networks tackles constrained optimization of an objective function, where the constraints are those specifying the network. Tractable inference is based on enumeration procedures in the style of BNs. With respect to a given $K^{SE}(\mathbf{V})$, consider computation (by marginalization) of lower bound $\underline{P}(x_0)$, with X_0

[29] Remember we always assume $\nu < \infty$.

and x_0 *target* node and event, respectively:

$$\underline{P}(x_0) = \min_{\substack{P(X_i|pa(X_i)) \in K(X_i|pa(X_i)), \\ pa(X_i) \in \Omega_{Pa(X_i)}; i=0,\dots,n}} \sum_{(x_1,\dots,x_n) \in \Omega_{\mathbf{V}\setminus\{X_0\}}} \prod_{i=0}^{n} P(x_i|pa(X_i)) \,. \tag{1.10}$$

Eq. (1.10) requires optimization of a multilinear function over the feasible region, induced by the (linear) constraints on the elements of the network [69, 8]. Credal marginalization is NP-hard [70].

For a query event $(\mathbf{X}_Q = \mathbf{x}_Q)$, exact credal updating on $(\mathbf{E} = \mathbf{e})$ requires solving:

$$\underline{P}(\mathbf{x}_Q|\mathbf{e}) = \min_{P(X_i|pa(X_i)) \in K(X_i|pa(X_i)); i=0,\dots,n} \frac{\sum_{\mathbf{x}_{-(Q,\mathbf{E})} \sim (\mathbf{x}_Q, \mathbf{e})} \prod_{i=0}^{n} P(x_i|pa(X_i))}{\sum_{\mathbf{x}_{-\mathbf{E}} \sim \mathbf{e}} \prod_{i=0}^{n} P(x_i|pa(X_i))} \,. \tag{1.11}$$

Eq. (1.11) generalizes Eq. (1.6). Analogously to Eq. (1.10), it corresponds to a constrained optimization problem, targeting the minimum (maximum) value for $P(\mathbf{x}_Q|\mathbf{e})$. Constraints follow from i) the Markov condition, yielding the right-hand side of Eq. (1.6), ii) consistency of the CPTs with their associated CSs in the model, that is $P(X_i|pa(X)) \in K(X_i|pa(X_i))$, for some fixed $pa(X_i) \in \Omega_{Pa(X_i)}$, for any $X_i \in \mathbf{V}$. Standard belief updating for CNs is NP-complete on singly connected networks [176], with the notable exceptions of polytrees where all nodes are binary [99], general DAGs with $\mathbf{E} \equiv \mathbf{V}\setminus\mathbf{X}_Q$ and tree-augmented classifiers [101]; it is NP^{PP}-complete on general DAGs.

Updating under strong independence subsumes MAP inference in BNs via the Cano-Cano-Moral (CCM) transformation [30]. This is achieved by introduction of auxiliary vacuous root nodes. Each is a parent of a credal node, indexing its extreme points Instantiation of auxiliary nodes to some configuration induces a BN: MAP is solved following standard routines (cfr Sec. 1.3). MAP in BNs may be symmetrically be solved by credal belief updating [69]. Additionally, solution of the MPE problem has polynomial complexity on singly connected networks, otherwise it is NP-complete.

Combinatorial optimization techniques as well as mathematical programming procedures may be considered for inference with CNs.[30] Additionally, several algorithms for exact inference with CNs, based on local search, have been proposed in the literature [30, 60, 70, 72]. Approximate algorithms have also been proposed for locally specified CNs [9, 30, 60, 177]. Exact routines proved inefficient on general DAGs[31], while, among approximate inference techniques, none managed to find a satisfactory

[30] See, respectively [99, 9, 31, 177] and [69, 227].

[31] Following [8], consider any binary node X, with $|Pa(X)| = k$, and suppose every $Y \in Pa(X)$ is defined on a h-valued sample space. Up to h^k optimization tasks need solving, with the space of all possible solution having 2^{h^k} candidates with respect to single node X. Such task is inefficient even for the simple case of $k = 4, h = 3$!

trade-off between efficiency [9, 30, 60] and accuracy [177]. See [15, Sec. 9.5.3] for a survey on the algorithms for belief updating with CNs.
As a general remark, when given a CN, pre-processing of the DAG ought always be considered, to improve efficiency of inference. This involves: i) removal of all (unobserved) leaf nodes, ii) dropping arcs from nodes in $\mathbf{E}$, iii) replacement of nodes in $\mathbf{X}_Q$ and in $\mathbf{E}$ by binary variables, taking values either consistent with the input to the inferential task, that is with query event and with $\mathbf{e}$, or its negation.
We now introduce *ApproxLP*, an efficient algorithm proposed by Antonucci for approximate inference with constraint-based specified CNs [8]. ApproxLP, or ApproxLP, solves a sequence of (compact) linear problems, improving previous existing techniques [59]. It is well suited to CNs with exponentially large sets of extreme points. Also, it serves as an exact procedure in some special settings.

The ApproxLP Algorithm

Let us consider Eq. (1.10). The sum-product term on the right hand-side of Eq. (1.10) may be equivalently specified as:

$$P(x_0) = \sum_{x_j, pa(X_j)} P(x_0|x_j, pa(X_j))P(pa(X_j))P(x_j|pa(X_j)) \tag{1.12}$$

The third term on the right hand-side is readily available from the network specification, whereas the remainders can be obtained by conditioning. Whenever conditioning event $(x_j, pa(X_j))$ has zero probability, we require the second and third terms of the right hand-side of the equation to floor the product down to zero [8]. This way, such term has null probability and does not (safely) appear in objective function. ApproxLP tackles upper (lower) approximation of $\underline{P}(x_0)$ ($\overline{P}(x_0)$), by iteratively fixing all local CSs but one to singletons. Let X_j, $j = 1, \ldots, n$, be any fixed *free* node at iteration $t = 0$, one PMF $P_t(X_i|pa(X_i)) \in extK(X_i|pa(X_i))$ is chosen, for every $pa(X_i) \in \Omega_{Pa(X_i)}$, for all $i \neq j, i = 1, \ldots, n$. The resulting CN is a near-Bayesian network (near-BN), with a single credal node, say X_j. By Eq. (1.12), Eq. (1.10) is reduced to:

$$\underline{P}_t(x_0) = \min_{P(X_j|pa(X_j)) \in extK(X_j|pa(X_j))} \sum_{\substack{x_j \in \Omega_{X_j}, \\ pa(X_j) \in \Omega_{Pa(X_j)}}} P_t(x_0|x_j, pa(X_j))P_t(pa(X_j))P(x_j|pa(X_j))$$

This way, optimization is restricted to the set of extreme points of $K(X_j|pa(X_j))$, for every $pa(X_j) \in \Omega_{Pa(X_j)}$. This follows from the linearity of the program[32]. We expect

[32] The solution of a linear program is known to lie on an extreme point of the feasible region, of $K(X_j|pa(X_j))$, in our case.

$\underline{P}_t(x_0) \geq \underline{P}(x_0)$, since restriction of Eq. (1.10) to a near-BN implies introduction of additional constraints, potentially reducing the feasible region.
Iteration $t+1$ picks another *free*, say $X_{j'}$ and fixes $P_{t+1}(X_j|pa(X_j))$ to be the lower bound for $(X_j|pa(X_j))$, obtained by solving the linear optimization task, for every $pa(X_j)$. The procedure is iterated until convergence. Similar reasoning applies $\overline{P}(x_0)$; let $\overline{P}^*(x_0)$ be the value resulting from the program, we expect ApproxLP to incur in an internal approximation: $\overline{P}^*(x_0) \leq \overline{P}(x_0)$.

Some remarks are due. First, if the CN is a near-BN, ApproxLP provides the exact solution to the marginalization task, after a single iteration. This observation extends to the case of k mutually marginal independent nodes, the exact solution requiring k iterations. Also, inference may be readily extended to, e.g. belief updating (the program goes from linear to linear-fractional) [8]. As a third point, consider initial step $t = 0$; choice of $P_0(x_i|pa(X_i))$, for every x_i and every $pa(X_i)$, $i = 1, \ldots, n$, may consist in either: i) sample at random an extreme point from $K(X_i|pa(X_i))$, ii) take the pignistic transform [231] of $K(X_i|pa(X_i))$, i.e. its center of mass. This latter option aims at preventing possibly misleading choices for P_0, and applies to the case of exponentially large sets of extreme points. Yet, since we expect the solution not to include any inner points, each random variable in $\mathbf{V}\setminus\{X_0\}$ must be *freed* at least once; i.e. n iterations are required, at least. As a final remark, for a fixed CN, let $h = \max_{i=0,\ldots,n} |\Omega_{X_i}|$ and d be the treewidth of a DAG (as from Sec. 1.3. Also, let q be the maximum number of linear constraints that define a CS. Eq. (1.10) has at most h^{d+1} random variables and qh^d constraints. ApproxLP requires time equivalent to that required to run a linear programming solver on the input specification, whose size is proportional to qh^d, at each iteration [8].

Different irrelevance concepts may be considered, including EI and ER. Since we will only consider CNs specified under strong independence, a detailed discussion on other extensions for a CN is out of the scope of this work. We refer the interested reader to [52, 54, 71, 76] for details.

This chapter introduced basic concepts for standard and imprecise probabilities. DAG-based models were also introduced, with a focus on the issue of independence in their specification. Ch. 2 will consider probabilistic generalized belief revision whereas *graphical belief revision* will be characterized in details throughout Ch. 3 to Ch. 5, where most concepts from the current chapter will play a key role.

Chapter 2

Probabilistic Belief Revision

Belief change theory studies the way a doxastic agent adjusts her belief upon new information in a static context [134]. First contribution to the theory date back to Harper's paper in 1975 [126], and were followed by Levi's philosophical work on the subject in 1980 [157]. However, it is Alchourrón, Gärdenfors and Makinson's seminal paper in 1985 [3] that ended up being broadly recognized as the landmark for the theory, also referred to as AGM paradigm after its founders.

By AGM theory, the belief set of a doxastic agent (or You) is defined by a logically closed set of (finitely many) sentences in a propositional language $\mathcal{L}$ (Sec. 2.1). Sentences represent beliefs held by the agent, provided $\mathcal{L}$ is subject to an ordering induced by a Tarskian logic [242].[1][2] An agent's epistemic state contains both a belief set, and a dynamic component that governs changes of belief states [133, Ch.17].

Throughout the present work, we will consider what Williamson refers to in his book as *external propositional logic* [263, p.121]. Probability functions are ascribed to propositions in $\mathcal{L}$, and inference consists in reasoning under uncertainty.[3] Properly, in our setup an agent's epistemic state contains a belief set whose dynamic component is a probability measure.[4] With subjective probabilities, this approach further specializes to Bayesianism [263, 94], and changes in the agent's beliefs are aimed to satisfy new distributional constraints [84, 92]. We shall use term *probabilistic belief revision*, to avoid misinterpretations, when referring to a doxastic agent's epistemic state that reflects a quantitative attitude toward her degree of credence

[1] See [214, p.129] and [133, Ch.17] for details.

[2] Belief sets are called *theories* when they are logically closed with respect to such consequence operator; that is, when they may be intended as sets of Your *doxastic commitments* [158].

[3] See also [103].

[4] Other dynamic components might be considered, see [235, 118, 133].

for a given proposition, rather than a qualitative judgment.[5]

Belief revision operators, or rules, are functionals that combine prior knowledge, i.e. Your deductively closed set of accepted beliefs, with externally provided pieces of information. This latter may result from an observational process, possibly affected by degrees of uncertainty, or may come as externally provided beliefs on the world outside (e.g. up-to-date experts' statements). We characterize fully reliable observations on a single case as probabilistic instances, and call them *specific knowledge*, as opposed to *generic knowledge*, this latter informing the agent on the whole system [93].

A general introduction to belief revision is provided in Sec. 2.1, using the language of classical propositional logic. Probabilistic belief revision with imprecise probabilities is discussed in Sec. 2.2. Sec. 2.3 introduces a new class of adjustment operators, based on Lewis' Imaging rule, tackling the case of *inconsistent* generalized instances.
One-shot belief revision [92], i.e. in a static setting, only is considered in the present chapter. This adjusts an agent's belief on a static system when a single probabilistic instance is available. Implementation of revision rules from this chapter for belief propagation with DAG-based models (BNs and CNs) will be considered in Ch. 3. A thorough discussion on belief aggregation, or *opinion pooling*, may be found in Ch. 4, accounting for multiple overlapping instances. Finally, the iterated case is a major focus in Ch. 5, when evidence is available on a collection of non-overlapping elements of a domain.

2.1 Introduction to Probabilistic Belief Revision

Let Ω be any space of atoms - atomic (boolean) propositional variables - and let a *world* ω be an assignment of truth to each element from Ω, such that there exist up to $2^{|\Omega|}$ *conceivable worlds*. $\mathcal{L}$ is the set of all *propositional formulae* over Ω. Any propositional formula $\phi \in \mathcal{L}$ is satisfied by the worlds in $[\phi] \subseteq \Omega$. Formally, ω satisfies ϕ writes $\omega \models \phi$; such that $\omega \in [\phi]$ if and only if $\omega \models \phi$. Logical connectives $\{\wedge, \vee, \neg, \rightarrow, \leftrightarrow\}$ may be used to concatenate several formulae. Finally, $\top$ and $\bot$ denote, respectively, tautology and contradiction.
As already mentioned, a doxastic agent is equipped with a deductively closed set of propositions, or belief states, represented by K, her *belief set* over *possible worlds*

[5] In the general propositional setting, a further distinction must be operated upon full beliefs and probability-one events. This is out of the scope of our work; we refer the reader to [133, Ch.17] for a soft introduction to the subject, and related issues.

$A \subseteq \Omega$. Elements of K are closed sets of formulas, in the propositional logic language $\mathcal{L}$ governed by the Tarki's consequence operator.[6]
When given a belief set, three main operations are relevant to adjust it and satisfy any available formula ϕ: contraction, expansion and revision [3]. Roughly, suppose You believe $\neg\phi$ is true but You are apprised ϕ is actually to be believed. Adjustment of Your belief set must occur while preserving overall consistency. First, $\neg\phi$ must be safely removed by contraction: not only $\neg\phi$ must be canceled from the set, but all sentences it entails must be also considered, to avoid inconsistencies. At this point, introduction of ϕ in Your belief set, by expansion, is straightforward. Combination of contraction and expansion is usually referred to as revision; inter-reducibility of revision and contraction was proved by the well-known Levi's and Harper's identities.
AGM postulates, best known in their KM formulation of Katsuno and Mendelzon [147], present two major shortcomings when belief revision involves i) conditional formulae, and ii) in the iterated setting. We refer the reader to Ch. 5 for a discussion on the latter.
Belief revision operators are related to non-monotonic logic.[7] This is in turn related to probability theory via the so-called *ϵ-probability functions*, i.e. PMFs ranging in the probability simplex, augmented by infinitesimal values $\epsilon > 0$. ϵ-probability functions also bridge belief revision operators to pooling functionals for belief merging, that will be considered by Ch. 4.[8]

In our probabilistic framework, a belief state over A shall be represented by PMF P_A:

$$P_A(A) = \{P(\omega) : \omega \in A, P(\omega) \geq 0, \sum_{\omega \in A} P(\omega) = 1\} .$$

Similarly, P_Ω is defined with respect to every $\omega \in \Omega$. We shall just write P, when domain is clear from the context.
Let $\mathbf{V}$ be a collection of $(n+1)$ discrete r.v.s, $n \geq 0$. There, $\omega \equiv \mathbf{v}$ and $\Omega \equiv \Omega_{\mathbf{V}}$, with $|\Omega| < \infty$. The collection of all possible probabilistic statements on Ω reduces to $\{\phi \bowtie c : \phi \in \mathcal{L}, \bowtie \in \{=, \geq, \leq, <, >\}, c \in [0,1]\}$. Also, A reduces to any arbitrary tautology, such that P_A is strictly positive on A, equal to zero otherwise. For a given formula ϕ,

$$P([\phi]) = \sum_{\mathbf{v} \in \Omega_{\mathbf{V}} : \mathbf{v} \sim A} P(\mathbf{v}) \mathbb{I}_{\mathbf{v} \models \phi} ,$$

[6]An alternative representation to total pre-ordering of possible worlds, or *entrenchment ordering*, by an agent's belief makes use of Grove's system of spheres [119].
[7]Properly, to *rational consequence relations* for consistency preservation [167].
[8]See [262] and references therein for details.

with $\sim$ denoting consistency among atomic events. e.g. let $n = 3$, $\phi = \{x_1 \wedge x_2\}$, $[\phi] = \{(x_0, x_1, x_2) : x_0 \in \Omega_{X_0}\}$.
In the general case, probabilistic beliefs are specified by a set of linear constraints, to define a CS K. Let Φ denote a collection of formulae, $K^{\Phi} \subseteq K$ is the subset of beliefs that satisfy Φ. Any belief state P satisfies Φ, i.e. $P \models \Phi$, whenever it satisfies each $\phi \in \Phi$. Any set Φ is accepted whenever it is consistent with each $P \in K$, it is rejected if its negation only, $\neg\Phi$, is, or it is neutral if both are consistent. Let $c \in [0, 1]$, for a given formula ϕ, $P \models (\phi \bowtie c)$ whenever $P(\phi) \left(= \sum_{\mathbf{v}:\mathbf{v}\sim[\phi]} P(\mathbf{v})\right) \bowtie c$.
Let P be any model, assigning values to events $\alpha \in \Sigma$. We account for new pieces of information as probabilistic constraints on the behavior of P over some partition $\Sigma' \subseteq \Sigma$, e.g. via another PMF P' on coarse domain Σ'. See [215, Table 1] for a general overview in the propositional framework.
We define any adjustment rule (AR) $\circ$ as a functional that combines P and $P'_{\Sigma'}$, denoted as P', forcing the first to accommodate for the second. Formally, let α be any target event, $(P \circ P')(\alpha) = P^{\circ}(\alpha)$ such that $d(P^{\circ,\downarrow\Sigma'}, P') \leq \delta$; $d(\cdot, \cdot)$ and δ being any fixed, respectively, distance measure and arbitrary small value.
AR $\circ$ is *responsive* [89] to P' whenever it holds

$$(P \circ P')(\sigma') = P^{\circ}(\sigma') = P'(\sigma'),$$

σ' being any atom[9] in Σ'. Equivalently, $d(P^{\circ,\downarrow\Sigma'}, P') = 0$. Any responsive AR is also known as obeying the *success* postulate [165], or performing *full meet revision* [163].

As a second principle for belief adjustment, we require our model to be as close as possible to its prior states of knowledge; this is known as *minimal change principle* [25, 132]. In words, we want P° to retain all prior knowledge on events that are modeled independently of partition Σ'. Three types of interactions may occur among *adjusting* and *adjusted* events [89], that we briefly outline. Let α be any event in Σ, and P° be the model, as revised by P':

Full Irrelevance $P(\alpha) = P^{\circ}(\alpha)$,

Weak Relevance $P(\alpha) = kP^{\circ}(\alpha)$, $k > 0$,

Relevance $\nexists k > 0$ such that $P(\alpha) = kP^{\circ}(\alpha)$.

Following Dietrich [89], two main approaches to probabilistic belief revision may be considered: conservative or Distance-based. The first is characterized in its strong

[9] This may be generalized to the case of coarse partitions of Σ', i.e. to general events $\alpha' \subseteq \Sigma'$.

or weak forms, depending on P' being, respectively, strongly or weakly silent on events $\alpha \in \Sigma$. Any $\circ$ satisfies *Conservativeness* whenever it holds:

$$P^{\circ}(\alpha|\sigma') = P(\alpha|\sigma') \quad \forall \alpha \in \Sigma, \forall \sigma' \in \Sigma' . \tag{2.1}$$

AR $\circ$ satisfies the *minimal distance* principle whenever it holds:

$$P^{\circ} = \operatorname{argmin}_{\tilde{P} \in \mathcal{P}'} d(P, \tilde{P}) , \tag{2.2}$$

with fixed $d(\cdot, \cdot)$ and $\mathcal{P}' \subseteq \mathcal{P}$ being the set of all models on Σ consistent with P'. Among others, the Kullback Liebler (KL) divergence [154] was used to measure discrepancy between a PMF P_{Σ} and its revision P°. The KL divergence is defined as follows:

$$KL(P^{\circ} \| P) = \sum_{\omega \in \Omega} P^{\circ}(\omega) \log \frac{P^{\circ}(\omega)}{P(\omega)} . \tag{2.3}$$

Eq. 2.3 is not a proper distance measure, as it fails symmetry. It is easy to see $KL(P^{\circ} \| P)$ is always non-negative, and reaches zero if and only if $P^{\circ} = P$. Also, the argument of the logarithmic term, on the right hand-side of Eq. 2.3, reduces to ratio $P^{\circ}(\sigma')/P(\sigma')$, for each $\sigma' \in \Sigma'$, under Conservativeness. A shortcoming of KL divergence as benchmark for Distance-based probabilistic belief revision was pointed out by Chan and Darwiche in their influential paper of 2005 [34]. There, a toy example was used to prove Eq. 2.3 might floor toward zero in some extreme cases, while (very) small probability values are in fact being revised by a large amount, such that the relative change is potentially (very) large. The authors proposed using the CD-distance [34] instead, defined as:

$$CD(P^{\circ}, P) = \max_{\omega \in \Omega} \log \frac{P^{\circ}(\omega)}{P(\omega)} - \min_{\omega \in \Omega} \log \frac{P^{\circ}(\omega)}{P(\omega)} = \log \frac{\max_{\omega} \frac{P^{\circ}(\omega)}{P(\omega)}}{\min_{\omega} \frac{P^{\circ}(\omega)}{P(\omega)}} .$$

Also in this case, it may be easily proved ratios in $CD(P^{\circ}, P)$ reduce to $P^{\circ}(\sigma')/P(\sigma')$, for each $\sigma' \in \Sigma'$, under Conservativeness.

Several alternative systematizations of ARs were proposed to Dietrich *et al.*'s, and no general consensus was reached so far. e.g. both Conservativeness and Minimal Distance are expressed by a general *Minimal Change* axiom in [165]. This may be due to the fact ARs tackling Eq. (2.1) or (2.2) yield equivalent solutions in standard settings. Also, Conservativeness preserves prior knowledge, and thus serves as a viable solution to Distance-based tasks; the converse is not true, in general. Indeed, we argue distinction into either Conservative (or Kinematical) or Distance-based approaches ought to be accounted for, as the epistemic standpoints differ from each other. The upcoming sections specialize discussion to Your knowledge base being represented by a joint CS (cfr Ch. 1), while some observational (or informational) process requires adjustment of the model. Probabilistic evidence on some random variable X is called *inconsistent* when it contradicts certainty (or impossibility) in Your belief base. We refer to revision and adjustment, respectively, to distinguish the case of consistent evidence from the general setup.

2.2 Probabilistic Belief Revision

In the probabilistic framework, Your epistemic state is represented by joint CS $K(\mathbf{V})$ over Ω. Let ϕ be any upcoming formula, postulates KM1-KM6 in [147] translate as follows:

KM1 $(K \circ \phi) \models \phi$,

KM2 Let $K \models \phi$, $(K \circ \phi) \equiv (K \cup \phi)$,

KM3 If $\phi \neq \perp$, then $(K \circ \phi) \neq \perp$,

KM4 If $K_1 \equiv K_2$ and $\phi_1 \equiv \phi_2$, then $(K_1 \circ \phi_1) \equiv (K_2 \circ \phi_2)$,

KM5 If $(K \circ \phi) \models \psi$, then $(K \circ (\phi \wedge \psi))$, for any further formula ψ,

KM6 If $(K \circ \phi) \models \psi$, then $(K \circ (\phi \wedge \psi))$ implies $((K \circ \phi) \models \psi)$.

For a given $X \in \mathbf{V}$, we define marginal probabilistic evidence by PMF P'_X over Ω_X[10], such that $P(x) \neq P'_X(x)$ for at least one $x \in \Omega_X$. We call $P'_X(X)$ *soft evidence* (SE) following Valtorta *et al.* [244]. SE bears an impression of the degree of reliability that is associated to each (forecasted) event from Ω_X, i.e. on the *evidence of uncertainty* [197]. It may be also intended as a set of probabilistic constraints on the system modeled by P [58]. Furthermore, it is easy to see SE on r.v. X shall be equivalently expressed by formula $\phi_x = (\{x\} = c_x)$, such that $P'_X(X) = \Phi_X = \{\phi_x : \sum_{x \in \Omega_X} c_x = 1\}$.
By *Partiality* [50], revision of P by P'_X requires preservation of zero-probability events. Rationality of such principle was advocated by several authors, e.g. [85, 262]. Simply put, an agent's belief ought to be calibrated with available evidence. If this is so, certainty on the occurrence of an event in Ω_X, say x', floors $P(x)$ to zero, for every $x \neq x'$(cfr Kolmogorov axioms, Ch. 1). If the agent accepted to change her mind on x in light of new evidence, then she would rather be reasonably sure, rather than certain about X being x'; but then $P(x') < 1$, and $P(x) > 0$ for at least one $x \neq x'$. Hence, Partiality requires certainty on the occurrence of x' must

[10]Remember we assumed X is a discrete r.v. over countable possibility space Ω_X.

imply certainty in P'_X as well (and symmetrically with zero-probability events). As already mentioned, we will refer throughout to partial operators as *revision rules*, as opposed to general ARs.

Part of the contributions in the remainder of this section may be also found in [171].

2.2.1 Belief revision with Soft Evidence

We refer to *kinematical mechanics* for the adjustment of a belief set as to consistency principles for Conservativeness, that we are willing to favor over a mere Distance-based approach [25].[11] We introduce *probability kinematics* following Wagner's characterization [254].

Definition 11 (Probability kinematics [139, 254]). *Let P and P° be any two PMFs over (Ω, Σ), and let Ω_X be a countable collection of pairwise disjoint events in Σ, i.e. a coarse partition of $\Omega(\equiv \Omega_{\mathbf{V}})$. P° comes from P on Ω_X based on probability kinematics (PK) if there exists a sequence $P'_X(X) = \{P'_X(x) : x \in \Omega_X, \sum_{x\in\Omega_X} P'_X(x) = 1\}$ such that it holds:*

PK1 $P^\circ(\alpha|x) = P(\alpha|x)$*, for each $x \in \Omega_X$, and for any event $\alpha \in \Sigma$, (Conservativeness)*

PK2 $P^\circ(X) = P'_X(X)$*. (Responsiveness)*

By PK, P is changed to agree with P'_X (PK2), while preserving relevance of each $x \in \Omega_X$ to any event $\alpha \in \Sigma$ (PK1). It is straightforward to see PK1 is just Eq. (2.1). Jeffrey's Rule below comes naturally as an equivalent definition of PK [84]:

Definition 12 (Jeffrey's Rule [139]). *Let P, P° and P'_X as above. Probabilistic Belief Revision operator Jeffrey's Rule ($\circ_J$) adjusts P to comply with P'_X as follows:*

$$(P \circ_J {P'}_X)(\alpha) = \sum_{x\in\Omega_X} P(\alpha, x)\frac{{P'}_X(x)}{P(x)} .$$

We denote the Jeffrey Revision of P by some SE as P°_J}, and require $0/0 = 1$.

See [254] on a further PK-based revision rule.

It may be readily seen, deterministic knowledge of event $(X = x)$ shall be equiva-

[11] See [215, Sec.3] on foundational/coherent and vertical/horizontal perspectives in the symbolic framework.

lently specified by $P'_X(x) = 1$, 0 otherwise.[12], whence:

$$\left(P \circ_J P'_X\right)(\alpha) \equiv P(\alpha|x)\,, \tag{2.4}$$

where the righ hand-side is just conditioning. Such *hard* evidence [244] trivially corresponds to ($\phi = \{x\}$) in a propositional language. Eq. (2.4) above shall serve as a further principle for any kinematical rule, requiring reducibility of any $\circ$ to conditioning, when SE strengthens to hard.

2.2.2 Belief revision with Conditional Soft Evidence

Suppose SE is gathered conditional on another r.v. Y taking value $y \in \Omega_Y$; or, without loss of generality, on $(\mathbf{Y} = \mathbf{y})$, $\mathbf{Y} \subseteq \mathbf{V}\backslash\{X\}$, $\mathbf{y} \in \Omega_{\mathbf{Y}}$. We define conditional SE (CoSE) as the collection of probabilistic statements:

$$P_{X|y}(X|y) = \left\{ P'_{X|y}(x|y) : P'_{X|y}(x|y) \geq 0 \wedge \sum_{x \in \Omega_X} P'_{X|y}(x|y) = 1, (x, y) \in \Omega_X \times \{Y = y\} \right\}.$$

CoSE may be equivalently specified by propositional $\Phi_{X|y}$, set of formulae, whose generic element is $\phi_{x|y} = (\{y \to x\} = c_x)$, provided $\sum_{x \in \Omega_X} c_x = 1$.

Definition 13 (Conditional PK [27])**.** *Let P and P° be any two PMFs on (Ω, Σ). Let $P(y) > 0$, P° comes from P on $\Omega_X \times \{Y = y\}$ based on conditional PK (CoPK) if there exists a sequence $P'_{X|y}(X|y)$ as above such that it holds:*

CoPK1 $P^\circ(\alpha|x, y) = P(\alpha|x, y)$, *for each $x \in \Omega_X$, for any $\alpha \in \Sigma$, (Conditional Conservativeness)*

CoPK2 $P^\circ(\alpha|y') = P(\alpha|y')$, *for each $y' \in \Omega_Y\backslash\{y\}$ and any $\alpha \in \Sigma$, (Irrelevance of Neutral Conditioning Events)*

CoPK3 $P^\circ(Y) = P(Y)$, *(Irrelevance to Conditioning Events)*

CoPK4 $P^\circ(X|y) = P'_{X|y}(X|y)$. *(Conditional Responsiveness)*

The following revision rule extends Def. 12 to the conditional setting:

[12] As a remark, while probabilistic findings extend standard evidence, they do not necessarily result from an observation process. e.g. they may be gathered as forecasts produced by external sourced whose system of knowledge is not disclosed - betting odds -, or qualitative evaluations from experts. Thorough characterization of uncertain evidence may be found in the survey of [188], and related works. There, probabilistic evidence is further distinguished into fixed and not-fixed. Such distinction will be critical to iterated belief revision, in Ch. 5.

Definition 14 (Adams Conditioning [27, 90])**.** *Let P, P° and $P'_{X|y}$ as above, with $P(y)>0$. CoPK-based [27, Th.5] probabilistic belief revision operator $\circ_A$ yields the Adams Revision $\left(P^{\circ_A}_{X|y}\right)$ of P, consistent with $P'_{X|y}$, if it is obtained as:*

$$\left(P \circ_A P'_{X|y}\right)(\alpha) = P(\alpha, \neg y) + \sum_{x \in \Omega_X} P(\alpha, x, y) \frac{P'_{X|y}(x|y)}{P(x|y)},$$

provided $0/0 = 1$.

It is easy to see $\circ_A$ reduces to $\circ_J$ if $P(y) = 1$[13] - and, trivially, $P^{\circ_A}_{X|y}(\alpha) \equiv P(\alpha|x,y)$ whenever $(X,Y) = (x,y)$ is fully observed. Revising mechanics of a PMF on CoSE that are not based on CoPK, such as those based on Minimal-Entropy[14], yield severe inconsistencies in the general case. Among others, Pearl [194] advocated its usage fails to adequately handle causal evidence, while being computationally intractable. Also, Friedman and Shimony [104] showed the updates they yield are not reducible to simple conditioning, *ceteris paribus*, whereas Seidenfeld [221, Ch.3] criticized the lack of independence on the agent's language of such probabilities. See additionally [120, 27, 90].
We introduce an illustrative toy example to motivate generalization of probabilistic belief revision to the conditional setting[15].

Example 7 (The Professor's Disease)**.** *Rumors have persistently reported Professor A contracted a serious infectious disease ($A = a$), while on summer vacation. B and C are both students of Prof. A. They attend different classes and do not know each other, although they are both close friends of E, who is not a student. The described setup shall be represented by the BN in Fig. 2.1 (analogous to the one from Fig. 1.3), and the following CPTs:*

A	**P(A)**
a	0.75

,

B	**A**	**P(B|A)**
b	a	0.60
b	$\neg\, a$	0.05

,

C	**A**	**P(C|A)**
c	a	0.60
c	$\neg\, a$	0.01

and

E	**B**	**C**	**P(E|B,C)**
d	b	c	0.90
d	b	$\neg\, c$	0.75
d	$\neg\, b$	c	0.55
d	$\neg\, b$	$\neg\, c$	0.05

[13] I.e. $y = A = \top$, $A \subseteq \Omega$.

[14] Enhanced by, e.g.[263, 2].

[15] The reader familiar with the literature of artificial intelligence might notice our example is analogous to the so-called *Judy Benjamin* problem [120], inspired by the 1980 movie *Private Benjamin*.

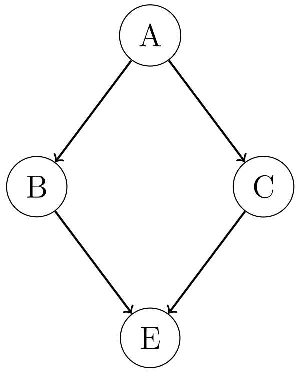


Figure 2.1: A toy multiply connected Bayesian network, with $\mathbf{V} = \{A, B, C, E\}$.

E's doctor has to decide wether prescribing her patient a test for the disease. She is interested in knowing how likely it is for E to have contracted the infection from either B or C, provided that her beliefs about Prof. A being infected in the first place are unchanged. She does not have any information about anyone's health status among A, B and C; she only knows E spends more time with B than with C, as from the CPTs above.
Based on her current knowledge, the doctor believes there is 50.5% *chances that E is now ill. As the test is invasive, the doctor asks E to give her additional information, to better discriminate between the choice of prescribing it or not. All that E reports is that C reached Prof. A after class to ask questions about the course. According to the doctor's knowledge of the transmissibility of the disease, if Prof. A was infectious, chances of C getting the disease are now increased from 60% to 75%. By straightforward application of Adams conditioning (see Def. 14), the revised probability of E being infected ($E = e$) grows to* 53.8%*; she is more prone to prescribe the test.*

2.2.3 Belief revision with Credal Soft Evidence

Just like Your beliefs may be encoded by a CS K on Ω, probabilistic evidence may come as a (closed and convex) collection of PMFs K'_X on Ω_X. We call K'_X *credal* SE (CSE), extending sharp probabilistic evidence to the case $|extK'_X(X)| \geq 1$:

$$K'_X(X) = \left\{ P'_X(x) : \underline{P}'_X(x) \leq P'_X(x) \leq \overline{P}'_X(x), x \in \Omega_X \right\} .$$

K'_X might be equivalently specified by the collection of formulae Φ_X, with generic element $\phi_x = (\{x\} \bowtie c_x)$, $c_x \in [0,1]$, provided $\sum_{x\in\Omega_X} c_x \bowtie 1$, $\bowtie \in \{=, \leq, \geq\}$. [16]

[16]Also, to guarantee $P'_X(x) \in [0,1]$, we require $P'_X(x) \leq 0$ and $P'_X(x) \geq 1$ always reduce to equalities.

We distinguish *precise* from *credal* probabilistic belief revision as follows: the first refers to a sharp belief set being revised by single-valued uncertain instances, while the second tackles the generalized case of credal beliefs revised by either sharp or imprecise assessments.
Set-valued instances may be readily provided, e.g. following a qualitative assessment, or rather result from the convexification of a collection of sharp assessments gathered on the same possibility space. In this latter case, we resort to belief merging, that we characterize throughout Ch. 4. By the frequentist approach, an observational process may produce set-valued findings when records are incomplete. We refer in this case to a *incompleteness process* [271], producing observations on a coarse grid of (ordered) Ω_X. A *missingness mechanism*, on the other hand, prevents some elements of a system from observation upstream, yielding a *blank* observation. A simplistic, although quite involved [121], assumption is that of *coarsening at random* (CAR) [128, 117], or *missing at random* (MAR) [164], respectively.
By the CAR assumption, the mechanism preventing precision is non-informative. Let $P(\alpha|\mathbf{E}=\mathbf{e},\mathbf{M}=*)$ be the probability of event α, with $(\mathbf{E}=\mathbf{e},\mathbf{M}=*)$ being a given finding; random variables in $\mathbf{E}$ are those actually observed, while those in $\mathbf{M}$ are missing. If CAR is assumed, $P(\alpha|\mathbf{E}=\mathbf{e},\mathbf{M}=*)=P(\alpha|\mathbf{E}=\mathbf{e})$. A naïve updating rule would project the dataset to its complete component [78].
Indeed, non-selective missingness mechanisms are unrealistic in many cases, and there are degrees of ignorance on the dynamics leading to incomplete records; see [270] on *mixed knowledge*. The *conservative updating rule* (CUR) was proposed in this direction by De Cooman and Zaffalon [78] for a reliable treatment of missing data, to update beliefs under a near-ignorance belief state on the missingness process. This was successively extended by the conservative inference rule (CIR) [270] for mixed missingness mechanisms, yielding imprecise revisions of any belief set. As previously argued, this approach also fits into an *objectivist* perspective: imprecise modeling of a PMF into a CS is motivated by (partial) ignorance about data, particularly about the missingness mechanisms. Such a frequentist interpretation may be justified by the connection, established by [10], between (graphical) credal updating and CIR (see Sec. 3.2.2). CIR is defined as follows:

$$\underline{P}(\alpha\|^{\mathbf{M}_r}\mathbf{E}=\mathbf{e},\mathbf{M}_{nr}=*)=\min_{\mathbf{m}_{nr}\in\Omega_{\mathbf{M}_{nr}}}\underline{P}(\alpha|\mathbf{e},\mathbf{m}_{nr}) \tag{2.5}$$

where $\mathbf{M}_r$ are random variables that are MAR (and are thus canceled out), $\mathbf{E}$ are actually observed, $\mathbf{M}_{nr}$ are missing due to unknown mechanisms; $\mathbf{M}$ denotes all missing random variables.
CUR is just Eq. (2.5) when $\mathbf{M}_r=\emptyset$, and hence:

$$\underline{P}(\alpha|\mathbf{E}=\mathbf{e},\mathbf{M}_{nr}=*)=\min_{\mathbf{m}_{nr}\in\Omega'_{\mathbf{M}_{nr}}}\underline{P}(\alpha|\mathbf{e},\mathbf{m}_{nr})\,, \tag{2.6}$$

with $\Omega'_{\mathbf{M}_{nr}} \subseteq \Omega_{\mathbf{M}_{nr}}$ induced by partial (coarse) observations. A similar approach was proposed in [207] for parameter learning with BNs.[17]

We hereby extend Def. 11:

Definition 15 (Credal PK). *Let Ω, $\mathbf{V}$ and X as before. Also, let K and K° be any joint CSs over $\mathbf{V}$. K° comes from K on Ω_X based on credal probability kinematics (CPK) if there exists a collection $K'_X(X) = \{P'_X(x) : \underline{P}'_X(x) \leq P'_X(x) \leq \overline{P}'_X(x), x \in \Omega_X, \sum_{x\in\Omega_X} P'_X(x) = 1\}$ such that, for any event $\alpha \in \Sigma$:*

CPK1 $K^\circ(\alpha|x) = K(\alpha|x)$*, for each $x \in \Omega_X$, (Credal Conservativeness)*

CPK2 $K^\circ(X) \supseteq K'_X(X)$*. (Credal Responsiveness)*

Based on [276] and previous discussion, a further principle may be included:

$$K^\circ(\alpha) = K(\alpha|x)\,,$$

when $K'_X(X)$ degenerates to hard evidence on $(X = x)$, that we refer to as CPK3, bridging credal updating and revision operators.
Extension of Jeffrey's rule to the imprecise setting requires simultaneous computation of all bounds spanned by the updating of all SEs, based on PMFs consistent with the CS:

Definition 16 (Credal Jeffrey's Rule). *Let K be any CS over (Ω, Σ), and suppose credal SE K'_X is provided. $K_X^{\circ_{CJ}}$ is called the Credal Jeffrey Revision of K on K'_X if, for any event $\alpha \in \Sigma$, it holds:*

$$K_X^{\circ_{CJ}}(\alpha) = \left\{ P_X(\alpha) : P_X(\alpha) = (P \circ_J P'_X)\,(\alpha),\ \begin{array}{l} P \in K, \\ P'_X \in K'_X \end{array} \right\}.$$

The lower envelope of the Credal Jeffrey Revision of P writes:

$$\underline{P}_X^{\circ_{CJ}}(\alpha) = \min_{\substack{P\in K \\ P'\in K'}} \sum_{x\in\Omega_X} P(\alpha|x)P'_X(x)\,. \tag{2.7}$$

Thus, in the general case $|extK| > 1$ minimization is no longer a linear task. If $|extK| = 1$, Credal Jeffrey's rule requires simultaneous revision of the prior by all PMFs in K'_X.

[17] A further revision setup may be considered when dealing with imprecise probabilities. There, revision corresponds to construction of a present and of a future bet, whose combination pursues sure loss avoidance in time. See [272] for a thorough discussion and characterization of the subject.

Proposition 1. *Let K° and K be two joint CSs over the same domain Ω, the first being the Credal Jeffrey Revision of the second on $K'_X(X)$. Let K^* be a further CS, resulting by $\circ_{CJ}$ restricted to $extK(\mathbf{V})$ and $extK'_X(X)$. It holds:*

$$K^\circ \equiv K^* .$$

Proof. Proof is trivial and follows from the characterization of a CS as the convexification of its own extreme points. □

By Prop. 1, $K_X^{\circ_{CJ}}$ may be efficiently computed by taking the convex hull of all Jeffrey Revisions resulting from all possible combinations of the extreme points. A similar result was proved by [211] with respect to another class of ARs, that will be discussed in Sec. 2.3.
We prove the following consistency results for precise belief revision based on CSE:

Theorem 1. *Let P and K° be a joint PMF and CS, respectively, on $\Omega_{\mathbf{V}}$, and let $K'_X(X)$ be CSE over $X \in \mathbf{V}$. If K° results from P by $\circ_{CJ}$, then the revision process is based on CPK, and CPK2 is strongly satisfied, i.e. inclusion strengthens to equality.*

Proof. Without loss of generality, let $\alpha = \{y\}$ and $\mathbf{V} = \{X, Y\}$. Since $K^\circ(\mathbf{V})$ is obtained by Credal Jeffrey's rule, CPK1 holds by definition. Also, it holds:

$$\begin{aligned}
\underline{P}^\circ(x) &= \min_{P^\circ \in K^\circ} \sum_{y \in \Omega_Y} P^\circ(y, x) \\
&= \min_{P^\circ \in K^\circ} \sum_{y \in \Omega_Y} P(y|x) P^\circ(x) \\
&= \min_{P'_X \in K'_X} P'_X(x) \sum_{y} P(y|x) \\
&= \underline{P}'_X(x) ,
\end{aligned}$$

and analogously for upper bound $\overline{P}^\circ(x)$, for each $x \in \Omega_X$. CPK2 is strongly satisfied, since $K^\circ(X) = K'(X)$. □

The orthogonal setup - revision of a credal belief set by SE - was prevously considered by Da Rocha *et al.* [58]. In their contribution, SE poses probabilistic constraints, that are used to adjust a given K by multilinear programming. Particularly, K specifies a Credal network, revised by SE that is properly transformed into virtual instances. We will return to this in Ch. 3.

Th. 2 extends Th. 1 above:

Theorem 2. *Let K and K° be two joint CSs, and let $K'_X(X)$ be some CSE over $X \in \mathbf{V}$. If K° results from K by $\circ_{CJ}$, then the revision process is based on CPK, and CPK2 is strongly satisfied.*

Proof. Without loss of generality, let $\alpha = \{y\}$, $\mathbf{V} = \{X, Y\}$. As for CPK2, it holds by definition (and by Prop. 1):

$$\begin{aligned}
\underline{P}^\circ(x) &= \min_{P^\circ \in K^\circ} \sum_{y \in \Omega_Y} P^\circ(y, x) \\
&= \min_{\substack{P \in extK \\ P'_X \in extK'_X}} P'_X(x) \sum_{y \in \Omega_Y} P(y|x) \\
&= \min_{P'_X \in K'_X} P'_X(x) \\
&= \underline{P}'_X(x)
\end{aligned}$$

and analogously for upper bound $\overline{P}^\circ(x)$, for each $x \in \Omega_X$. This way, $K^\circ(X) = K'_X(X)$.
Proof CPK1 is satisfied is trivial from Th. 1 and Prop. 1. □

We lay bare kinematical principles for imprecise probabilistic belief revision of K by (sharp) CoSE $P'_{X|y}$ on $\Omega_X \times \{Y = y\}$:

Definition 17 (Credal Conditional PK)**.** *Let K and Ω_X as above, and let K° be any other CS over $\mathbf{V}$. K° comes from K on $\Omega_X \times \{Y = y\}$ based on credal conditional probability kinematics (CCoPK) if there exists some CoSE $P'_{X|y}$ such that it holds:*

CCoPK1 *$K^\circ(\alpha|x, y) = K(\alpha|x, y)$, for each $x \in \Omega_X$, (Credal Conditional Conservativeness)*

CCoPK2 *$K^\circ(\alpha|y') = K(\alpha|y')$, for each $y' \in \Omega_Y \backslash \{y\}$, (Credal Irrelevance of Neutral Conditioning Events)*

CCoPK3 *$K^\circ(y) = K(y)$, for each $y \in \Omega_Y$, (Credal Irrelevance to Conditioning Events)*

CCoPK4 *$K^\circ(X|y) \ni P'_{X|y}(X|y)$, or equivalently $K^\circ(X|y) \models \Phi_{X|y}$. (Generalized Responsiveness)*

As a remark, CCoPK trivially reduce to CoPK when $|extK(\mathbf{V})| = 1$. Once again, we introduce the further requirement of reducibility of imprecise probabilistic belief revision upon CoSE to credal updating.
In the general case, i.e. when $|extK(\mathbf{V})| \geq 1$, the following operator may be considered:

Definition 18 (Credal Adams Conditioning)**.** *Let K, K° and CoSE $P'_{X|y}$ as above, for some $y \in \Omega_Y$, $Y \in \mathbf{V}$, provided $\underline{P}(y) > 0$. Revision operator $\circ_{cA}$ yields the Credal Adams Revision $\left(K^{\circ_{cA}}_{X|y}\right)$ of K consistent with $P'_{X|y}$:*

$$\left(K \circ_{cA} P'_{X|y}\right)(\alpha) = \left\{P^{\circ_A}_X(\alpha) : P^{\circ_A}_X(\alpha) = \left(P \circ_A P'_{X|y}\right)(\alpha), P \in K\right\} .$$

It holds:

Theorem 3. *Credal Adams conditioning is based on Credal CoPK.*

Proof. Without loss of generality let $\mathbf{V} = \{X, Y, Z\}$, with $\alpha = \{z\}$ and $\mathbf{Y} = \{Y\}$. Let K° be the joint CS obtained by Credal Adams conditioning, by Def. 18, the lower envelope of $K(z)$ writes:

$$\underline{P}^{\circ}(z) = \min_{P \in K^{\circ}} \left[P(z, \neg y) + \sum_x P(z, x, y) \frac{P'_{X|y}(x|y)}{P(x|y)}\right] . \tag{2.8}$$

Proof of Credal Irrelevance of Neutral Conditioning Events (CCoPK2) follows trivially from Eq. (2.8), that is reduced to the first term only on the right hand-side. Analogously, Credal Irrelevance to Conditioning Events (CCoPK) is straightforward for any $y' \neq y$, while $\overline{P}(y)$ follows from the conjugacy relation. $\overline{P}(\neg y)$ and $\underline{P}(y)$ are derived symmetrically.
Based on Eq. (2.8), we prove Credal Conservativeness is satisfied as follows:

$$\begin{aligned}
\underline{P}^{\circ}(z|x, y) &= \min_{P \in K^{\circ}} \frac{P(z, x, y)}{P(x, y)} \\
&= \min_{P \in K} \frac{P(z, y|x) P'_{X|y}(x|y)}{P(y) P'_{X|y}(x|y)} \\
&= \min_{P \in K} P(z|x, y) \,,
\end{aligned}$$

and analogously for the upper case.
As for Generalized Responsiveness, inclusion of $P'_X(X|y)$ by $K^{\circ}(X|y)$ follows from Th. 2. □

Extension of Credal Adams conditioning to the case of credal CoSE would require weakening of kinematical principles, due to detrimental dilation - loose inclusion relationships - that is expected to result from the revision process. Beyond that, we do not stick to the details of such a task, as revision of a CS by a conditional credal instance foreshadows a quite involved discussion on its actual usefulness and reliability.

All revision rules presented so far assume *preservation of certain events* [109] (when consistent evidence is used to revise Your belief), and Partiality. With credal probabilistic belief revision, as a consequence of the two, a further principle of *Preservation of vacuous knowledge* also results (as proved by [200]). In the upcoming section we resort to Imaging [159] as a rule for belief adjustment upon (possibly) inconsistent pieces of information.

2.3 Relaxing Partiality: Imaging Operators

Contributions from this section may be also found in [170].

From previous considerations, we refer to inconsistencies as to failures of the Partiality principle, with a major focus on zero-probability events. *Imaging* was introduced by Lewis in 1976 [159], as a non-trivial alternative to conditioning on such inconsistent events. Roughly, it represents the "*thought experiment by a minimal action*" [106] that makes a formula consistent.
Going back to the propositional language, if some world ω is inconsistent with formula ϕ, according to a knowledge base, Imaging shifts beliefs towards those elements of Ω that are closest to ϕ, called ϕ-worlds. $\gamma(\omega, \phi)$ is a *closest world function*, mapping ω to its closest ϕ-world; see [160] for a detailed discussion. In our formalism, $(\phi = \{x\})$ yields $\gamma(\mathbf{v}, \phi) = (\mathbf{v}\backslash\{X\}, x) \in \Omega$, for any $\mathbf{v} \in \Omega$.

Definition 19 (Imaging [159])**.** *Let P be any PMF over (Ω, Σ). For a given ϕ and closest world function $\gamma(\cdot, \phi)$. $P_\phi^{\circ_I}$ is the Image of P on ϕ if it is obtained by Imaging operator $\circ_I$ as:*

$$(P \circ_I \{\phi\})(\alpha) = \sum_{\omega' \in \alpha} \sum_{\omega \in \Omega} P(\omega) \mathbb{I}_{\gamma(\omega,\phi)=\omega'} .$$

In Lewis' words, by Imaging on event ϕ, *"Probability is moved around, but not created or destroyed"*, while *"Every share stays as close to it as it can to the world it was originally created"* [159, pp. 310-311]. To summarize: i) inconsistent evidence is accounted for in the Image of P, whereas conditioning is left undefined; ii) Imaging changes the whole belief set to comply with reliable knowledge ϕ, while conditioning re-defines the domain of P, focusing on those worlds in Ω consistent with ϕ.[18]

Example 8. *Let $\mathbf{V} = \{X, Y\}$, with $P(x, Y) = 0$, $P(\neg x, y) = 0.6$ and $P(\neg x, \neg y) = 0.4$. Imaging on $(\phi = \{x\})$ yields $(P \circ_I \{X = x\})(y) = 0.6$, which corresponds to $P(y)$. If conditioning was applied, $P(Y|x)$ would not be defined.*
Consider $\alpha = \{x\}$, $(P \circ_I \{X = x\})(x) = 1$: $\circ_I$ adjusts P to always be consistent with $\phi = \{x\}$.

[18] See [160] for details.

Generalized forms of Imaging were introduced in the literature, see, e.g. [109, 211]. See also [277] on a unifying approach to belief adjustment.

In a recent paper Günther [123] introduced Jeffrey Imaging, that we denote as $\circ_{jI}$, for the generalized case of probabilistic formula $(\phi = c)$, with $c \in [0,1]$.[19] Adjustment operator $\circ_{jI}$ trivially extends Ramachandran *et al.*'s Partial Imaging [206].

Definition 20 (Jeffrey Imaging [123, 206])**.** *Let P be any PMF over (Ω, Σ). For a given formula $\{\phi = c\}$, with $c \in [0,1]$, $P_X^{\circ_{jI}}$ comes from P by Jeffrey Imaging $\circ_{jI}$ on $\{\phi = c\}$ if it holds:*

$$(P \circ_{jI} \{\phi = c\})(\alpha) = P_\phi^{\circ_I}(\alpha)c + P_{\neg\phi}^{\circ_I}(\alpha)(1-c)$$

We denote the Jeffrey Image of P on $\{\phi = c\}$ as $P_\phi^{\circ_{jI}}$.

Both Imaging and Jeffrey Imaging are *homomorphic* change functions[20], i.e. they define a structure-preserving map. A generalized characterization of Jeffrey Imaging will be provided below, within the multivalued imprecise probabilistic framework (cfr Def. 23).

In this section we tackle probabilistic belief adjustment by (possibly inconsistent) sharp or imprecise probabilities, following an approach based on the Imaginary counterparts of PK. See in this spirit [165, 276] on belief functions (and previous section on CSE).
Following [276] (and previous section), we are willing to check a further consistency requirement, that would reproduce Eq. (2.4). This way, any adjustment *kinematical* operator reduces to some form of conditioning when probabilistic evidence strengthens to full observation.

With probabilities, evidence on some random variable X is *inconsistent* when it contradicts certainty (or impossibility) in Your knowledge base, a collection of deductively closed propositions, i.e. belief states, represented by CS K over $\mathbf{V}$. We provide an example to motivate our contribution.

Example 9. *Celeste is swimming in a lake and sees some black birds from the distance. She knows black birds living around that lake are rather tame, while swans are very aggressive. Also, she is sure only white or grey swans exist, although those she sees actually look like swans. As she is reasoning, she is apprised by a sailor that a small group of black swans has been spotted around that area. Should she be worried about the birds she sees?*

[19] Günther's definition assumes $c \in (0,1)$.
[20] See [109] and [206, Obs.1], respectively.

Classic belief revision operators from Sec. 2.2 fail to revise a belief set based on information from an observational process when inconsistencies of this kind arise. This feature was motivated in the literature by the already mentioned principle of Partiality. Roughly, Your belief ought to be calibrated with the evidence already available, if any. Although rationality of Partiality has been advocated by several authors [85, 262], an AR for the adjustment of a model to any piece of evidence ought to be in a doxastic agent's toolbox, to avoid building a new model from scratch when unexpected information shows up.[21] Such an operator updates the knowledge base to be consistent with new evidence, while leaving previous beliefs on related events as unchanged as possible. We characterize optimality requirements for such adjustment operators as *Imaginary Kinematics*, and extend them to deal with generalized forms of evidence. Analogously to previous section, we consider probabilistic evidence, generalized to conditional assessments, and to imprecise ones. Again, those latter may be intended as originating from a qualitative judgment. We introduce adjustment functionals based on Lewis' Imaging, and study their features and properties.

We lay bare the kinematical conditions that ought to be satisfied by any AR, when (possibly) inconsistent probabilistic evidence is gathered.[22]
Consider general probabilistic evidence on r.v. X: K'_X on Ω_X, such that $|\Omega_X| \geq 2$; equivalently, $\Phi_X = \{\phi_x, \sum_{x\in\Omega_X} c_x = 1\}$. We introduce Imaginary Kinematics as the counterparts of PK (and CPK) to Imaging:

Definition 21 (Imaginary Kinematics)**.** *Any joint CS K° on $\mathbf{V}$ comes from K by imaginary kinematics (IK) on (possibly inconsistent) credal evidence K'_X on r.v. X whenever it holds:*

IK1 $K^\circ(\alpha|x) \equiv K_x^{\circ_I}(\alpha)$*, for any $\alpha \in \Sigma$ and each $x \in \Omega_X$,*

IK2 $K^\circ(X) \models \Phi_X$*, for each $x \in \Omega_X$,*

IK3 $K^\circ(X) \equiv K_x^{\circ_I}(X)$ *whenever $c_x = 1$ for some $x \in \Omega_X$.*

Analogously, based on Def. 17, we provide an imaginary characterization of CPK:

Definition 22 (Imaginary Conditional Kinematics)**.** *Let K, K° as above, such that $\underline{P}(y) > 0$ for each $y \in \Omega_Y$. K° comes from K on $\Omega_X \times \{Y = y\}$ based on imaginary conditional kinematics (ICK) if there exists a (possibly inconsistent) sequence $P'_{X|y}$ (or collection $\Phi_{X|y}$) such that it holds:*

[21] As a remark, we consider inconsistencies referring to events from a partition of Ω that have zero (upper) probability.

[22] With imprecise probabilities, inconsistency occurs when $\overline{P}(x) = 0$, and positive evidence is provided on x, for some $x \in \Omega_X$.

ICK1 $K^{\circ}(\alpha|x,y) \models K^{\circ_I}_x(\alpha|y)$, *for each* $x \in \Omega_X$,

ICK2 $K^{\circ}(\alpha|y') \equiv K(\alpha|y')$, *for each* $y' \in \Omega_Y \backslash \{y\}$,

ICK3 $K^{\circ}(Y) \equiv K(Y)$,

ICK4 $K^{\circ}(X|y) \models \Phi_{X|y}$,

ICK5 $K^{\circ}(X|y) \equiv K^{\circ_I}_x(X)$, *whenever* $c_x = 1$, *for some* $x \in \Omega_X$.

For any $\alpha \in \Sigma$, if a CS K over $\mathbf{V}$ is used to represent Your beliefs, Imaging on $(\phi = \{x\})$ extends to:

$$(K \circ_I \{x\})(\alpha) = \{P^{\circ_I}_x(\alpha) = (P \circ_I \{x\})(\alpha), P \in K\} \,,$$

so that the lower envelope of K's Image on $\{x\}$, denoted as $K^{\circ_I}_x$, at α, writes:

$$\underline{P}^{\circ_I}_x(\alpha) = \min_{P(\mathbf{v}) \in K(\mathbf{V})} \sum_{\mathbf{v}' \sim \alpha} \sum_{\mathbf{v} \in \Omega_{\mathbf{V}}} P(\mathbf{v}) \mathbb{I}_{\gamma(\mathbf{v},x)=\mathbf{v}'} \,.$$

By [211, Th.1], $K^{\circ_I}_x$ may be efficiently obtained by taking the convex hull (CH) of the Image of each $P \in extK$ on $\{x\}$. Since the Image of each $P \in extK$ at $\alpha = \{x'\}$ trivially corresponds to $P^{\circ_I}_x(x') = 0$,[23] whenever $x' \neq x$, refinement of $K^{\circ_I}_x(X)$ degenerates to a single PMF satisfying $P'_X(x) = 1$, 0 otherwise. With an abuse of notation, this yields the following:

$$K^{\circ_I}_x(\mathbf{v}) \equiv \begin{cases} 1 \cdot K(\mathbf{v} \backslash \{X\}) & \mathbf{v} \sim x \,, \\ 0 & \text{otherwise} \end{cases} \,.$$

Example 10. *Let* K *over* $\mathbf{V} = \{X, Y\}$ *such that*

$$K \begin{pmatrix} x_1, y_1 \\ x_1, y_2 \\ x_2, y_1 \\ x_2, y_2 \\ x_3, y_1 \\ x_3, y_2 \end{pmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0.15 - 0.35 \\ 0.25 - 0.49 \\ 0 - 0.45 \\ 0.03 - 0.5 \end{bmatrix} .$$

It is easy to see $\underline{P}^{\circ_I}_{x_1}(y_j) = \underline{P}(y_j)$, $j = 1, 2$, *while* $\overline{P}^{\circ_I}_{x_1}(x_k) = 0$, $k = 2, 3$.

[23]By definition, $P^{\circ_I}_x(x) = \sum_{\mathbf{V} \in \Omega} P(\mathbf{V}) = 1$, 0 otherwise.

2.3.1 Imaging with Soft Evidence

We start from the case of sharp probabilistic evidence on Ω_X, i.e. $K'_X(X) = \{P'_X(X)\}$. The following adjustment operator extends Def. 20. As for Imaging above, notation that is used with sharp beliefs applies to the generalized case of belief sets, when $|extK| \geq 1$.

Definition 23 ((Probabilistic) Jeffrey Imaging)**.** *Let K be any joint CS over $\mathbf{V}$ as above. Suppose probabilistic evidence P'_X is provided over a (possibly) inconsistent collection of events, i.e. $\overline{P}(x) = 0$, whereas $P'_X(x) > 0$, for some $x \in \Omega_X$, $X \in \mathbf{V}$. For any event α, $K_X^{\circ_{jI}}$ is the Probabilistic Jeffrey Image of K if it holds:*

$$K_X^{\circ_{jI}}(\alpha) = \{P_X^{\circ_{jI}}(\alpha) = \sum_{x \in \Omega_X} P_x^{\circ_I}(\alpha) P'_X(x), P_x^{\circ_I} \in K_x^{\circ_I}, x \in \Omega_X\}\,.$$

That is, $K_X^{\circ_{jI}}(\alpha) = (K \circ_{jI} P'_X)(\alpha)$, for any $\alpha \in \Sigma$.

It holds:

Theorem 4. *Jeffrey Imaging is based on IK, and IK1 is strongly satisfied, i.e. $\models$ may be replaced by $\equiv$.*

Proof. To prove $\circ_{jI}$ is based on IK, we must check it produces a CS that satisfies IK1-IK3. Motivated by [211, Th.1], we restrict our attention toward the extreme points of K. Without loss of generality, let $\mathbf{X} = \{X, Y\}$. Each extreme point of $K(\mathbf{X})$, say $P_{j,k} \in extK$, may be equivalently specified as:

$$P_{j,k}(x, y) = P(x|x'_j)P(y|x, y'_k)\,, \tag{2.9}$$

with $P(y|x, y'_k)$ is set equal to zero whenever it is undefined and $P(x|x'_k) = 0$.[24] X' and Y' are uniformly distributed auxiliary random variables, used to index K's extreme points at X and at $Y|X$, respectively. This way, for a given ordering,

$$\begin{aligned} P(x|x'_1) &= \sum_{y'_k, y} P(x|x'_1)P(y|x, y'_k)P(y'_k) \\ &= \underline{P}(x)\,, \end{aligned}$$

and $\underline{P}(x, y) = P(x|x'_1)P(y|x, y'_1)$.
It holds:

$$\begin{aligned} \underline{P}_x^{\circ_I}(x) &= P_x^{\circ_I}(x|x'_1) \\ &= \sum_{y'_k, y, x} P(x|x'_1)P(y|x, y'_k)P(y'_k) \\ &= 1\,. \end{aligned}$$

[24] As a remark, $P(x) = 0$ does not necessarily imply $P(y|x) = 0$, in De Finetti's view.

If $P'_X \models (\phi = \{x\})$, IK3 is satisfied.
When a non-trivial PMF is provided, i.e. $P(x) > 0$ for at least two elements in Ω_X, it holds:

$$\begin{aligned} P_X^{\circ_j I}(x|x'_1) &= \left[\sum_{y'_k, y, x} P(x|x'_1) P(y|x, y'_k) P(y'_k) \right] P'_X(x) \\ &= 1 \cdot P'_X(x) \,, \end{aligned}$$

and similarly $P_X^{\circ_j I}(x|x'_{|extK|}) = P'_X(x)$. This proves IK2 since $K_X^{\circ_j I}(X) \ni P'_X(X)$.
Proof of IK1 is also straightforward:

$$\begin{aligned} P_X^{\circ_j I}(y|x, y'_1) &= \frac{\left[\sum_{x, x'_j} P(x|x'_j) P(y|x, y'_1) \right] P'_X(x)}{P'_X(x)} \\ &= \sum_{x, x'_j} P(x|x'_j) P(y|x, y'_1) \\ &= P_x^{\circ_I}(y|y'_1) \end{aligned}$$

Analogous reasoning applies to the upper envelope, and thus $K_X^{\circ_j I}(Y|x) \equiv K_x^{\circ_I}(Y)$.
This ends the proof. $\square$

Corollary 1. *Given sharp probabilistic knowledge on Ω_X, the Jeffrey Image of any CS may be equivalently specified by the convexification of all PMFs P°, each defined as follows:*

$$P^\circ(\alpha) = \sum_{\mathbf{v} \sim \alpha} P'_X(x) P(\mathbf{v}) \quad \forall P \in extK \,.$$

It is easy to see standard Imaging is also trivially based on IK.

Example 11. *Consider Ex. 10 above. Celeste's beliefs are formalized as follows: let $\Omega_Y = \{y \equiv Swan, \neg y \equiv \neg Swan\}$, $\Omega_X = \{x_W \equiv White, x_G \equiv Grey, x_B \equiv Black\}$ and $\Omega_Z = \{z \equiv Aggressive, \neg z \equiv Tame\}$.*
It holds:

$$P(Y) = \{(y, 0.7), (\neg y, 0.3)\} \;,$$

$$P(X|Y) = \left\{ \begin{array}{c} (x_W|y, 0.8), (x_G|y, 0.2), \\ (x_B|y, 0), (x_W|\neg y, 0.5), \\ (x_G|\neg y, 0.3), (x_B|\neg y, 0.2) \end{array} \right\} \,,$$

$$P(Z|Y) = \left\{ \begin{array}{c} (z|y, 0.95), (\neg z|y, 0.05), \\ (z|\neg y, 0.2), (\neg z|\neg y, 0.8) \end{array} \right\} \,.$$

According to Celeste's beliefs, $P(z|x_B) = 0.2$.
Based on the sailor's words, Celeste is willing to adjust her beliefs to be consistent with $P'_{X|y}(X|y) = \{(x_W, 0.8), (x_G, 0.1), (x_B, 0.1)\}$*. Straightforward application of Adams conditioning is undefined, since* $P(x_B|y) = 0$ *(probability of the event according to the original PMF), while CoSE* $P'_{X|y}(x_B|y) \neq 0$*. The same would occur with simple Jeffrey's rule, if any* $P'_X(x) \neq 0$ *was provided, given* $P(x) = 0$*, for some* $x \in \Omega_X$*. How could Celeste incorporate such reliable knowledge in her beliefs? and suppose* $P'_X(X) = \{(x_1, 0.3), (x_2, 0), (x_3, 0.7)\}$*. By Jeffrey Imaging on* P'_X*, we obtain* $K_X^{\circ_{jI}}(Y) \equiv K(Y)$*, while* $\underline{P}_X^{\circ_{jI}}(y_j|x_i) \equiv K_{x_i}^{\circ_I}(y_j)$*,* $i = 1, 2, 3$*,* $j = 1, 2$*. Also,* $K_X^{\circ_{jI}}(X) \models P'_X(X)$*, and* $K_X^{\circ_{jI}}$ *is equivalent to the convex hull of PMFs* P°*, defined as:*

$$P^\circ(x, y) = P'_X(x)P(y)\,,$$

for each $x \in \Omega_X, y \in \Omega_Y$ *and* $P \in extK$*.*

2.3.2 Imaging with Conditional Soft Evidence

We now introduce Adams Imaging as an adjustment operator $\circ_{aI}$, that extends $\circ_{jI}$ to the case of CoSE, just like revision rule $\circ_A$ extends $\circ_J$.

Definition 24 (Adams Imaging)**.** *Let* K *be any joint CS on* (Ω, Σ) *such that* $\underline{P}(y) > 0$*,* $Y \in \mathbf{V}$*, and let CoSE* $P'_{X|y}$ *on* $\Omega_X \times \{Y = y\}$*.* $K_{X|y}^{\circ_{aI}}$*, the Adams Image of* K *on* $P'_{X|y}$*, comes from* K *by Adams Imaging* $\circ_{aI}$*, if it holds:*

$$K_{X|y}^{\circ_{aI}}(\alpha) = \left\{ P_{X|y}^{\circ_{aI}}(\alpha) = P(\alpha, \neg y) + \sum_{x \in \Omega_X} P_x^{\circ_I}(\alpha, y)P'_{X|y}(x|y), \begin{array}{l} P \in K, P_x^{\circ_I} \in K_x^{\circ_I}, \\ x \in \Omega_X \end{array} \right\}.$$

I.e. $K_{X|y}^{\circ_{aI}}(\alpha) = \left(K \circ_{aI} P'_{X|y}\right)(\alpha)$*, for any* $\alpha \in \Sigma$*.*

When $|extK| = 1$, from previous considerations, Adams Imaging reduces to the following:

$$P_{X|y}^{\circ_{aI}}(\alpha) = P(\alpha, \neg y) + \sum_{x \in \Omega_X} P_x^{\circ_I}(\alpha, y)P'_{X|y}(x|y)\,. \tag{2.10}$$

Example 12 (Ex. 9 continued)**.** *The Adams Image on* $P'_{X|y}$ *of Celeste's beliefs on* $\Omega_X \times \Omega_Z$ *is the following:*

$$P_{X|y}^{\circ_{aI}}\begin{pmatrix} x_W z \\ x_W \neg z \\ x_G z \\ x_G \neg z \\ x_B z \\ x_B \neg z \end{pmatrix} = \begin{bmatrix} 0.5620 \\ 0.1480 \\ 0.0845 \\ 0.0755 \\ 0.0785 \\ 0.0515 \end{bmatrix}.$$

It holds $P^{\circ_{aI}}_{X|y}(X|y) = P'_{X|y}(X|y)$ *and* $P^{\circ_{aI}}_{X|y}(Y,Z) = P(Y,Z)$. *Adjustment of her beliefs by* $P'_{X|y}$ *yields* $P^{\circ_{aI}}_{X|y}(z|x_B) \approx 0.6$, *whereas* $P(z|x_B) = 0.2$. *Celeste rapidly swims back to shore.*

As a remark, inconsistency of $P'_{X|y}(x|y)$, for some $x \in \Omega_X$, with respect to any PMF P, may refer to either i) $P(x|y) = 0$, while $P(y) > 0$, (this is just the case of Adams Imaging above), or ii) $P(y) = 0$ in the first place, and possibly $P(x|y) = 0$. We argue case ii) deserves some caution, since full inconsistency of event $(Y = y)$ is likely not to yield any further conjecturing on related events, from a modeler's perspective. e.g. You are certain that no alien lives on Mars. Is it worth include Your belief on the alien having long hair in Your belief set, provided that You are not admitting the alien's existence upstream? On the other hand, we reckon arguments may be easily raised against our position, starting from our proposed running example. Still, so long no evidence is provided on Ω_Y, a cautious approach would require application of an iterated procedure. We leave this point for future work.

It is straightforward to see Adams Imaging generalizes Jeffrey's Imaging to the conditional setting, and thus, from previous results, Imaging. The following holds:

Theorem 5. *Adams Imaging is based on ICK, and ICK1 is strongly satisfied, i.e.* $\equiv$ *holds. Eq.* (2.10) *strongly satisfies all conditions.*

Proof. To prove $\circ_{aI}$ is based on ICK we need to check ICK1-ICK5 are satisfied by $K^\circ = \left(K \circ_{aI} P'_{X|y}\right)$. When $|extK| = 1$, ICK1-ICK5 reduce to the following:

ICK1' $P^\circ(\alpha|x,y) = P^{\circ_I}_x(\alpha|y)$, for each $x \in \Omega_X$,

ICK2' $P^\circ(\alpha|y') = P(\alpha|y')$,

ICK3' $P^\circ(Y) = P(Y)$,

ICK4' $P^\circ(X|y) = P'_{X|y}(X|y)$,

ICK5' $P^\circ(X|y) = P^{\circ_I}_x(X|y)$, whenever $P'_{X|y}(x|y) = 1$ for some $x \in \Omega_X$.

We first prove consistency points ICK4' and ICK5'. Let $P'_{X|y}$ be any PMF on $\Omega_X \times \{Y = y\}$, it holds:

$$\begin{aligned} P^{\circ_{aI}}_{X|y}(x|y) &= \frac{P^{\circ_I}_x(y) P'_{X|y}(x|y)}{\sum_x P^{\circ_I}_x(y) P'_{X|y}(x|y)} \\ &= P'_{X|y}(x|y) \end{aligned}$$

since $P_x^{\circ_I}(x,y) = P_x^{\circ_I}(y) = P(y)$, whatever $x \in \Omega_X$. Also, $\sum_x P'_{X|y}(x|y) = 1$ by definition. If $P'_{X|y}(x|y) = 1$ for some x, $P^{\circ_{aI}}_{X|y}(x|y) = 1$, 0 otherwise. The following holds:

$$\begin{aligned}\underline{P}^{\circ}(x|y) &= \min_{P^{\circ} \in extK^{\circ}} P^{\circ}(x|y) \\ &= P'_{X|y}(x|y)\frac{\overline{P}_x^{\circ_I}(y)}{\underline{P}_x^{\circ_I}(y)} \\ &\leq P'_{X|y}(x|y)\,.\end{aligned}$$

Similarly, $\overline{P}^{\circ}(X|y) \geq P'_{X|y}(X|y)$, for each $P^{\circ} \in extK^{\circ}$.
We now prove condition ICK1 (and thus ICK1') is satisfied by $\circ_{aI}$. Without loss of generality, let $\mathbf{X} = \{X, Y, Z\}$. It holds:

$$\begin{aligned}\underline{P}^{\circ}(z|x,y) &= \frac{P'_{X|y}(x|y)\underline{P}_x^{\circ_I}(z,y)}{P'_{X|y}(x|y)\overline{P}_x^{\circ_I}(y)} \\ &= \underline{P}_x^{\circ_I}(z|y)\,.\end{aligned}$$

As for point ICK2 (and ICK2'), it trivially holds by Def. 24:

$$\underline{P}^{\circ}(z|y') = \underline{P}(z|y')\,.$$

for any $y' \neq y$. ICK3' is proved analogously, since $P^{\circ_{aI}}_{X|y}(y) = 1 - P(\neg y) = 1 - \sum_{y' \neq y} P^{\circ_{aI}}_{X|y}(y')$. Similarly, fulfillment of ICK3 may be derived by the conjugacy relation [257]. □

Analogously to Cor. 1, it may be easily shown $K^{\circ_{aI}}_{X|y}$ at any $\mathbf{v} \sim y$ is equivalent to the CS obtained taking the product of sharp assessment $P'_{X|y}$ and the marginalization over r.v. X of the original belief set K. We provide the additional result, extending Rens *et al.*'s [211]:

Theorem 6. *Both Jeffrey and Adams Imaging satisfy consistency axioms KM1, KM3 and KM4. KM2, KM5 and KM6 are satisfied only is K is degenerate at $(X|y)$, i.e. $|K(X|y)| = 1$ (and at $(Z|w)$, for KM5 and KM6).*

Proof. Consider CS K and conditional probabilistic evidence $P'_{X|y}(X|y)$. To avoid cumbersome notation, we write $\circ$ to denote $\circ_{aI}$ throughout the proof. Also, we refer to general formula $\phi = c$ to denote both ϕ_x and $\phi_{x|y}$.
KM1 and KM3 follow from IK2 and ICK4 (cfr Th.1 and Th.2, respectively).
We prove KM2 is not satisfied under general conditions. Consider the lower envelope of K at $(x|y)$. If $K \models P'_{X|y}$, it holds:

$$\underline{P}(x|y) \leq P'_{X|y}(x|y)$$

by definition, and $\left(K \cup P'_{X|y}\right) = K$. From previous discussion, we expect $(K \circ P'_{X|y}) \supseteq K$, equality holding if and only if $K(\mathbf{X})$ may be equivalently specified as the product of sharp conditional assessment on $\Omega_X \times \{Y = y\}$ and CS over $(\mathbf{X}\backslash\{Y\}, y)$. Same reasoning applies to KM5 and KM6. These three postulates are satisfied if and only if K is already degenerate at the domain of probabilistic evidence, and consistent with it already.
Postulate KM4 holds by [211, Th.1]. $\square$

2.3.3 Imaging with Credal Soft Evidence

When beliefs are expressed as a joint CS over $\mathbf{V}$, adjustment by a single reliable PMF requires simultaneous computation of all bounds spanned by the updating of each $P \in K$. Also in this case, adjustment may be restricted to those PMFs in $extK$, and their convex hull (CH) efficiently considered.

Definition 25 (Credal Jeffrey Imaging)**.** *Given CS K over $\mathbf{V}$ and credal probabilistic evidence $K'_X(X)$, we define Credal Jeffrey Imaging $\circ_{cjI}$ as the functional mapping K to CS $K_X^{\circ_{cjI}}$, consistent with $K'_X(X)$ as follows:*

$$
\begin{aligned}
&K_X^{\circ_{cjI}}(\alpha) \\
&= \left\{ P^{\circ}(\alpha) = \left(P \circ_{jI} P'_X\right)(\alpha), \begin{array}{l} P(\mathbf{V}) \in K(\mathbf{V}), \\ P'_X \in K'_X(X) \end{array} \right\}
\end{aligned}
$$

The following preliminary result holds:

Lemma 1. *Let K be a joint CS over $\mathbf{V}$, and let K'_X denote a credal probabilistic finding, gathered on Ω_X. For any event α, the Jeffrey Image $K_X^{\circ_{cjI}}(\alpha)$ of $K(\alpha)$ on $K'_X(X)$ satisfies the following:*

$$K_X^{\circ_{cjI}}(\alpha) \supseteq K_X^{\circ_{cjI}}(\alpha|x) \supseteq K_x^{\circ_I}(\alpha)$$

for any $\alpha \in \Sigma$. Equality holds when $|K'(X)| = 1$.

Proof. Let $\mathbf{V} = \{X, Y\}$ and K be any CS over Ω. K'_X is gathered on Ω_X, to adjust K accordingly. By definition of Credal Jeffrey Imaging, it holds:

$$
\begin{aligned}
\min_{P_X^{\circ_{cjI}} \in K_X^{\circ_{jI}}} P_X^{\circ_{cjI}}(y|x) &= \min_{P(y) \in K(Y)} P(y) \frac{\underline{P}'_X(x)}{\overline{P}'_X(x)} \\
&\leq \min_{P(y) \in K(Y)} P(y)\,,
\end{aligned}
$$

and analogously for the upper envelope, with $\geq$. This proves the rightest inclusion relationship: $K_X^{\circ_{cjI}}(Y|x) \supseteq K_x^{\circ_I}(Y)(\equiv K(Y))$.
We now prove inclusion of $K_X^{\circ_{cjI}}(y|x)$ by $K_X^{\circ_{cjI}}(y)$:

$$
\begin{aligned}
\frac{\underline{P}_X^{\circ_{cjI}}(y)}{\underline{P}_X^{\circ_{cjI}}(y|x)} &= \frac{\underline{P}(y)\sum_x \underline{P}'_X(x)}{\underline{P}(y)\frac{\underline{P}'_X(x)}{\overline{P}'_X(x)}} \\
&= \overline{P}'_X(x) \sum_{x' \neq x} \underline{P}'_X(x') \\
&\leq 1 \,.
\end{aligned}
$$

Hence $\underline{P}_X^{\circ_{cjI}}(y) \leq \underline{P}_X^{\circ_{cjI}}(y|x)$, for any $x \in \Omega_X$, $y \in \Omega_Y$. $\overline{P}_X^{\circ_{cjI}}(y) \geq \overline{P}_X^{\circ_{cjI}}(y|x)$ is derived analogously.
Equality holds when $K'_X(X) = \{P'_X(X)\}$ as $\underline{P}'_X(x) = \overline{P}'_X(x)$, for each $x \in \Omega_X$, summing to one. □

The following result generalizes Th. 4 above:

Theorem 7. *Given (possibly) inconsistent credal probabilistic evidence, Credal Jeffrey Imaging yields the unique joint CS based on IK.*

Proof. Given a joint CS K over $\mathbf{V}$ and K'_X, let $\circ$ denote Credal Jeffrey Imaging. IK1 is satisfied by Lemma 1. IK2 is also satisfied as it holds:

$$\underline{P}_X^{\circ_{cjI}}(x) = 1 \cdot \underline{P}'_X(x)\,,$$

for each $x \in \Omega_X$. And analogously for $\overline{P}_X^{\circ_{cjI}}(X)$. When $K'_X(X) = \{P'_X(X)\}$ such that $P'_X(x) = 1$, IK3 is satisfied since $\circ_{cjI}$ reduces to $\circ_{jI}$. □

In this chapter we introduced kinematical adjustment operators in the generalized setting of imprecise probabilities, specified by credal sets. These are summarized in Table 2.1. Further generalization to the case of credal conditional probabilistic evidence is not straightforward as the adjustment process would likely incur in dilating mechanics, and hence yield detrimental loose inclusion relationships. This reasoning also applies to the iterated framework, where additional considerations must be formulated on the role evidence plays on the adjustment process. Future work will tackle this sort of scenarios.
Due to partiality, standard revision rules (conditioning, Jeffrey's rule and Adams conditioning) ought not be accounted for as fully general, whereas the remaining succeed in adjusting a given belief set following inconsistent observations. As for the special case of conditioning, an alternative approach worth mentioning was proposed in the literature of probabilistic beliefs: namely that of lexicographic beliefs

	Evidence (Φ_*)	**Partiality**	**Kinematics**
Conditioning ($\vert$)	$\{x\}$	Yes	[139, 84]
Jeffrey's Rule ($\circ_J$)	$\{x\} = c_x, \forall x$	Yes	[84]
Adams Conditioning ($\circ_A$)	$\{y \to x\} = c_x, \forall x$	Yes	[27, Th.5]
Credal Jeffrey's Rule ($\circ_{CJ}$)	$\{x\} \bowtie c_x, \forall x$	Yes	Th.1
CIR/CUR	$\{x\} \bowtie c_x, \forall x$	Yes	(see Sec. 3.2.2)
Lewis Imaging ($\circ_I$)	$\{x\}$	No	Th. 4[25]
Jeffrey Imaging ($\circ_{jI}$)	$\{x\} \bowtie c_x, \forall x$	No	Th. 4
Adams Imaging ($\circ_{aI}$)	$\{y \to x\} = c_x, \forall x$	No	Th.5
Credal Jeffrey Imaging ($\circ_{cjI}$)	$\{x\} \bowtie c_x, \forall x$	No	Th.7

Table 2.1: Summary of revision rules for probabilistic precise belief revision, introduced in Sec.2.2.

[24]. These are based on a radically different assumption that enriches the agent's beliefs by allowing for an internal hierarchy. Such approach eliminates the zero-probability events, and related issues, at their roots.

Chapter 3

Graphical Tools for Belief Propagation

In this chapter we introduce virtual evidence and extend it to the conditional and credal frameworks. Our approach stems from the seminal results of Chan and Darwiche [35], that proved inter-reducibility of virtual and soft instances. Such an equivalence result yields a convenient formalism for *graphical probabilistic belief revision*, although deep epistemic differences characterize virtual and soft evidence.[1] This is particularly relevant in the iterated framework, discussed in Ch. 5.
With graphical models, we intend probabilistic belief revision as *belief propagation* (or *focusing* [93], see also [92]), as opposed to *model revision*, i.e. elicitation of the model. Roughly, the first refers to *specific evidence* [93] on a case - e.g. the result of a diagnostic test on a patient - that is propagated by generalized forms of conditioning, to answer a given probabilistic query. *General evidence*, on the other hand is a statement on the world outside, aimed to replace some prior knowledge. We argue our methodology applies in principle to both tasks, with one-shot belief revision.

3.1 Related Work

Previous approaches to graphical probabilistic belief revision ([244, 197] and related works) tackle solution of a constrained optimization task, for a given distance measure. Minimization usually involves cross-entropy or total variation [197], yielding the I-projection [57] of the prior probability mass function by Iterative Proportional Fitting, i.e. iterated application of Jeffrey's rule on a subset $\mathbf{V}' \subseteq \mathbf{V}$ of

[1] See [188] and related works for a thorough discussion on this issue.

r.v.s. Entropy-based belief revision was proved to satisfy a number of principles [229]. Yet, consistency is likely to be failed within generalized contexts, as widely discussed, among others, by [120]. As previously discussed, a systematization of the techniques for probabilistic belief revision into either *distance-based* or *conservative* was proposed in [89]. The implications of a kinematical approach become more apparent with graphical models, where the pattern of independence plays a major role, and ought to be unchanged by any belief revision process. Simply put: kinematical approaches choose *conservativeness* over full *consistency*, and *vice versa* for those distance-based.

Belief revision of credal networks by sharp soft evidence was considered by [58], where the constrained optimization task was solved by multilinear programming, implementing variable elimination[2]. We deal with the opposite setting in Sec. 2.1 (and [171]), where we propose a set-valued quantification for reliable modeling of uncertain evidence in Bayesian networks.

Finally, Sec. 3.3 considers revision of a CN by credal probabilistic evidence. In this direction, although with a major focus on the merging process of credal sets, Adamcik [2] (and related works), provided several theoretical results. His approach is based on a class of distance-based operators (each minimizing a given Bregman divergence, see Ch. 4 for details). These, once again, are prone to fail standard postulates for revision operators, introduced in Sec. 2.1. Remarkably, uncertain belief propagation in a DAG-based model was also considered in the framework of evidence theory [230], where models specified by belief functions are revised by evidence affected by epistemic uncertainty. Yet, although a belief function can be regarded as a credal estimate, specialization of evidential rules to the graphical framework becomes more problematic and does not give a direct extension of the Bayesian networks formalism when DAGs are considered.

3.2 Generalized Pearl's Method with Bayesian Networks

Observation of random variable X may be affected by degrees of uncertainty, or be unreliable upstream.[3] Both situations yield evaluations of the elements of Ω_X based on likelihood ratios. Let D_X be the Boolean random variable representing the

[2]See p. 21 (Ch. 1).

[3]So far, we deliberately avoided any remarks about the link existing between a piece of information and its interpretation by the agent and/or its veridicity. See [148] and references therein for an introduction to the subject in the theory of belief revision.

observational process, we define *virtual evidence* (VE [194]) the collection

$$\lambda_X = \{\lambda_x : x \in \Omega_X\},$$

with $\lambda_x \propto P(D_X = d_X|X = x)$; d_X is called D_X's *truth* event, as opposed to its negation $\neg d_X$. VE is gathered to an agent for belief revision based on a "*nothing else considered*" approach [35] and, contrary to SE, it solely depends on the observational process itself. VE was first introduced by Pearl [194] within the framework of graphical models:

Definition 26 (Pearl's Method for VE [193]). *Let $\mathcal{B}$ be any BN. Given VE λ_X, augment $\mathcal{G}$ with auxiliary binary leaf node D_X, such that $Pa(D_X) = \{X\}$, with $\Omega_{D_X} = \{d_X, \neg d_X\}$. Its CPT is specified as:*

$$\begin{cases} P(d_X|x) \propto \lambda_x \\ P(\neg d_X|x) = 1 - P(d_X|x) \end{cases}$$

for each $x \in \Omega_X$. Let α be any target event, instantiation of D_X to its truth value yields:

$$P(\alpha|d_X) = \frac{\sum_{x\in\Omega_X} P(\alpha, x)P(d_X|x)}{\sum_{x\in\Omega_X} P(x)P(d_X|x)} \tag{3.1}$$

and we write, $P(\alpha|d_X) = (P \circ_P \lambda_X)(\alpha)$.

See Fig. 3.1(left panel) as an example.
It was proved VE may be transformed into SE, and *vice versa*:

Transformation 1 ([35]). *Let λ_X and P'_X be, respectively, VE and SE on random variable X. Let $P^{\lambda}_X(X)$ be defined as:*

$$P^{\lambda}_X(x) \propto \frac{P(x)\lambda_x}{\sum_x P(x)\lambda_x}.$$

for every $x \in \Omega_X$. Conversely, we define λ^P_X as:

$$\lambda^P_x = \frac{P'_X(x)}{P(x)}$$

for every $x \in \Omega_X$.

The following holds:

Proposition 2 ([35]). *Based on Tr. (1), it holds:*

$$(P \circ_P \lambda_X)(\alpha) = (P \circ_J P^{\lambda}_X)(\alpha) \tag{3.2}$$

$$(P \circ_J P'_X)(\alpha) = (P \circ_P \lambda^P_X)(\alpha) \tag{3.3}$$

This result was exploited, among others by [197], to generalize Def. 26:

Definition 27 (Generalized Pearl's Method for SE)**.** *Given $\mathcal{B}$ over $\mathbf{V}$, and $P'_X(X)$, with $X \in \mathbf{V}$, augment $\mathcal{G}$ with D_X as in Pearl's Method, and specify its CPT based on Tr. 1:*

$$\begin{cases} P(d_X|x) = \eta\lambda_x^P \\ P(\neg d_X|x) = 1 - P(d_X|x) \end{cases}$$

where η is any (strictly positive) real number that projects $P(d_X|x)$ in the probability simplex, shared by all columns of the CPT. Instantiate D_X to its truth value and propagate evidence, as in Eq. (3.1)*; $P^*(\alpha|d_X)$ results, for any fixed event α.*

Def. 27 implements $\circ_J$ by $\circ_P$ in a BN. This way, the revision process is reduced to an updating task. As a trivial consequence to Prop. 2, Pearl's RR $\circ_P$ for VE is based on probability kinematics [35, 255, 251], and soft information is fully retained by the network, i.e. $P(X|d_X) = P'_X(X)$. A further revision operator, that we call Wagner's rule, was proposed in [255], based on Bayes factors on Ω_X:

$$\kappa_X = \frac{P'_X(X)/P'_X(x_0)}{P(X)/P(x_0)} \propto \frac{\lambda_X}{\lambda_{x_0}},$$

x_0 being any reference value. Wagner's bridges Jeffrey's rule to Pearl's method as it tweaks the soft instances' connection to PMF P by focusing on changes in the odds. Both Wagner's rule and Pearl's method are Externally Bayesian, i.e. they commute with likelihood-based conditioning (see Ch 4). With VE, this results from commutativity of iterated updating. Let $P'_X(X)$ and $P'_Y(Y)$ be SE on r.v.s X and Y, respectively. Schema $X;Y$ corresponds to:

$$P^{X;Y}(\alpha) = \left(\left(P \circ_P \lambda_X^P\right) \circ \lambda_Y^{P^X}\right)(\alpha),$$

for any α. Revision of the PMF by P'_Y requires computation of denominator terms that differ from $P(Y)$, in general, as VE from Tr. 1 is not invariant to changes in the revision ordering. A simultaneous approach would require prior computation of all denominators, for all r.v.s considered.[4]
From a computational viewpoint, introduction of D_X does not alter the network's topology: complexity of inference is not affected by the procedure [193]. Yet, unlike straightforward application of Pearl's method, a preliminary inferential step is required to compute all denominator terms $P(X) = \{P(x) : x \in \Omega_X\}$. This step is trivially neglected if X is a root node.

[4]See [168] for a graphical justification of such procedure. We will come back to this in Ch. 5.

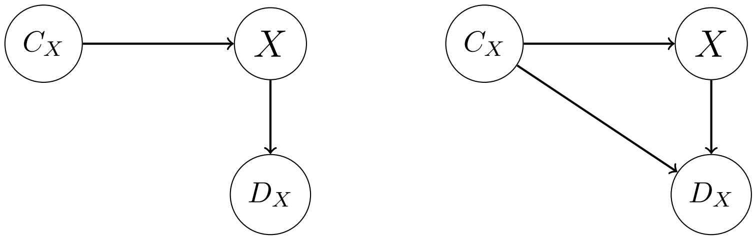


Figure 3.1: Graphical representation of generalized Pearl's methods on a toy Bayesian network when a marginal (left panel) or conditional (right panel) uncertain instance. CPTs of auxiliary binary leaf nodes D_X may be sharp or credal, based on the type of evidence.

Example 13 ([171]). *Let X denote the actual color of a traffic light with $\Omega_X := \{g, y, r\}$. Assume g (green) more probable than r (red), and y (yellow) impossible. Thus, for instance, $P(X) = [\frac{4}{5}, 0, \frac{1}{5}]$. We eventually revise $P(X)$ by SE $P'_X(X)$, which keeps yellow impossible, but assigns the same probability to the two other states, i.e. $P'_X(X) = [\frac{1}{2}, 0, \frac{1}{2}]$. Because of Eq. (3.3), this can be equivalently achieved by a VE $\lambda_X^P \propto \{1, 1, 4\}$. Vice versa, because of Eq. (3.2), VE $\tilde{\lambda}_X \propto \{1, 1, 5\}$ induces an updated $P_X^{\tilde{\lambda}}(X) = [\frac{4}{9}, 0, \frac{5}{9}]$. Such PMF coincides with $P(X|d_X)$ in a two-node BN, with d_X child of X, CPT $P(d_X|X)$ such that $P(d_X|X) = [\frac{1}{10}, \frac{1}{10}, \frac{1}{2}]$ and marginal PMF $P(X)$ as in the original specification.*

Uncertain evidence (UE) generalizes both concepts of SE and VE. Let $\mathbf{X}_U = \mathbf{X}_S \cup \mathbf{X}_V$ denote the collection of all random variables in $\mathbf{V}$ that are observed with uncertainty. We do not require $\mathbf{X}_S \cap \mathbf{X}_V = \emptyset$.

3.2.1 Generalized Pearl's Method with Conditional Uncertain Evidence

Let us now generalize CoSE to the case of uncertain conditional evidence, where $\lambda_{X|\mathbf{c}^*}$ is a collection of virtual instances on random variable $X \in \mathbf{X}_V$, conditional on relevant event $\mathbf{C}_X = \mathbf{c}^*_X$. To avoid cumbersome notation, we denote context variables as $\mathbf{C}_X = \mathbf{C}$. Conditional VE is defined as the collection:

$$\lambda_{X|\mathbf{C}=\mathbf{c}^*} = \{\lambda_{x|\mathbf{c}^*} : x \in \Omega_X\} \,. \tag{3.4}$$

Following previous reasoning, we define the following transformation:

Transformation 2. *Let $\lambda_{X|\mathbf{c}^*}$ and $P_{X|\mathbf{c}^*}$ be, respectively, CoVE and CoSE on random variable X, for a (single) given relevant context $\mathbf{c}^*$. CoSE $P^{\lambda}_{X|\mathbf{c}^*}(X|\mathbf{c}^*)$ is defined* *as:*

$$P^{\lambda}_{X|\mathbf{c}^*}(x|\mathbf{c}^*) \propto \frac{P(x|\mathbf{c}^*)\lambda_{x|\mathbf{c}^*}}{\sum_x P(x|\mathbf{c}^*)\lambda_{x|\mathbf{c}^*}} \quad \forall x \in \Omega_X \,.$$

Conversely, CoSE $P_{X|\mathbf{c}^*}(X)$ is transformed into CoVE $\lambda^P_{X|\mathbf{c}^*}$ as follows:

$$\lambda^P_{x|\mathbf{c}^*} := \frac{P_{X|\mathbf{c}^*}(x|\mathbf{c}^*)}{P(x|\mathbf{c}^*)} \quad \forall x \in \Omega_X \,.$$

Definition 28 (Pearl-Adams conditioning)**.** *Let conditional SE be available on X as $P'_{X|\mathbf{c}^*}(X|\mathbf{c}^*)$. By Pearl-Adams conditioning we augment the network with auxiliary binary leaf node $D_{X|\mathbf{c}^*}$ such that $Pa(D_{X|\mathbf{c}^*}) = \{X\} \cup \mathbf{C}$. Columns of node d_X's CPT consistent with relevant contexts are specified as from Tr. (2), with proportionality factor η:*

$$P(d_{X|\mathbf{c}^*}|x, \mathbf{c}^*) = \eta \frac{P'(x|\mathbf{c}^*)}{P(x|\mathbf{c}^*)} \quad \forall x \in \Omega_X \,. \tag{3.5}$$

If conditional VE is provided, they are specified as:

$$P(d_{X|\mathbf{c}^*}|x, \mathbf{c}^*) = \lambda_{x|\mathbf{c}^*} \quad \forall x \in \Omega_X \,. \tag{3.6}$$

All remaining columns are specified as:

$$P(d_X = d_{X|\mathbf{c}^*}|x, \mathbf{c}) = \eta \tag{3.7}$$

for every $\mathbf{c} \in \Omega_{\mathbf{C}} \backslash \{\mathbf{c}^*\}$.

Node $D_{X|\mathbf{c}^*}$ is instantiated to its truth *value and evidence is propagated. Given any target event α, $P^*(\alpha) = P(\alpha|d_{X|\mathbf{c}^*})$ results.*

See Fig. 3.1 (right panel) as an example. Unlike Pearl's method from Def. 26 and 27, the proposed augmentation of the network potentially affects the topology of the DAG as it introduces *multiple* paths, linking any node in $\mathbf{C}$ to auxiliary node $D_{X|\mathbf{c}^*}$. If $\mathcal{G}$ is singly connected, updating on virtual observation $(D_{X|\mathbf{c}^*} = d_{X|\mathbf{c}^*})$ takes no longer polynomial time and becomes an NP-hard task [45]. Several approaches to inference with multiply connected networks have been proposed in the literature, including inference based on junction trees (see Ch. 1). In Sec. 5.3 we propose application of the JT algorithm on iterated belief propagation; there, we require $|\mathbf{C}| < d$, treewidth of $\mathcal{G}$[5], in order not to affect the inference complexity.

Theorem 8. *Let CoSE be provided on the pair $(X, \mathbf{C})$ as $P'_{X|\mathbf{c}^*}(X|\mathbf{c}^*)$. Belief propagation by Pearl-Adams conditioning is based on CoPK.*

[5]Remember the treewidth of a DAG is defined as the maximum cardinality of all parents' sets, or equivalently, as the maximum in-degree.

Proof. We prove the PMF resulting from Pearl-Adams conditioning satisfies postulates i) to iv) of Def.17 when uncertain evidence is CoSE.
Proof of i) *Success*:

$$\begin{aligned}
P^*(x|\mathbf{c}^*) &= P(x|\mathbf{c}^*, d_{X|\mathbf{c}^*}) \\
&= \frac{P(\mathbf{c}^*)P(x|\mathbf{c}^*)P(d_{X|\mathbf{c}^*}|x,\mathbf{c}^*)}{P(\mathbf{c}^*)\sum_{x\in\Omega_X} P(x|\mathbf{c}^*)P(d_{X|\mathbf{c}^*}|x,\mathbf{c}^*)} \\
&= \frac{P(x|\mathbf{c}^*)P(d_{X|\mathbf{c}^*}|x,\mathbf{c}^*)}{\sum_{x\in\Omega_X} P(x|\mathbf{c}^*)P(d_{X|\mathbf{c}^*}|x,\mathbf{c}^*)} \\
&= P_{X|\mathbf{c}^*}(x|\mathbf{c}^*)\,.
\end{aligned}$$

Proof of ii) *Generalized rigidity*

$$\begin{aligned}
P^*(\alpha|x,\mathbf{c}) &= P(\alpha|x,\mathbf{c}, d_{X|\mathbf{c}^*}) \\
&= \frac{P(\alpha,x|\mathbf{c}, d_{X|\mathbf{c}^*})}{\sum_\alpha P(\alpha,x|\mathbf{c}, d_{X|\mathbf{c}^*})} \\
&= \frac{P(\alpha|x,\mathbf{c})P(x|\mathbf{c})P(d_{X|\mathbf{c}^*}|x,\mathbf{c})}{\sum_\alpha P(\alpha|x,\mathbf{c})P(x|\mathbf{c})P(d_{X|\mathbf{c}^*}|x,\mathbf{c})} \\
&= \begin{cases} \frac{P(\alpha|x,\mathbf{c})P_{X|\mathbf{c}^*}(x|\mathbf{c})}{\sum_\alpha P(\alpha|x,\mathbf{c})P_{X|\mathbf{c}^*}(x|\mathbf{c})} & \mathbf{c}=\mathbf{c}^* \\ \frac{P(\alpha|x,\mathbf{c})P(x|\mathbf{c})}{\sum_\alpha P(\alpha|x,\mathbf{c})P(x|\mathbf{c})} & \mathbf{c}\neq\mathbf{c}^* \end{cases} \\
&= P(\alpha|x,\mathbf{c})\,.
\end{aligned}$$

Proof of iii) *Rigidity of contexts*:

$$\begin{aligned}
P^*(\mathbf{c}) &= P(\mathbf{c}|d_{X|\mathbf{c}^*}) \\
&= \frac{\sum_{x\in\Omega_X} P(\mathbf{c})P(x|\mathbf{c})P(d_{X|\mathbf{c}^*}|x,\mathbf{c})}{\sum_{x\in\Omega_X,\mathbf{c}\in\Omega_\mathbf{C}} P(\mathbf{c})P(x|\mathbf{c})P(d_{X|\mathbf{c}^*}|x,\mathbf{c})} \\
&= \begin{cases} \frac{\sum_x P(\mathbf{c})P_{X|\mathbf{c}^*}(x|\mathbf{c})}{\sum_{x,\mathbf{c}} P(\mathbf{c})P_{X|\mathbf{c}^*}(x|\mathbf{c})} & \mathbf{c}=\mathbf{c}^* \\ \frac{\sum_x P(\mathbf{c})P(x|\mathbf{c})}{\sum_{x,\mathbf{c}} P(\mathbf{c})P(x|\mathbf{c})} & \mathbf{c}\neq\mathbf{c}^* \end{cases} \\
&= P(\mathbf{c})\,.
\end{aligned}$$

Proof of iv) *Rigidity to neutral contexts*:

$$\begin{aligned}
P^*(\alpha|\mathbf{c}) &= P(\alpha|\mathbf{c}, d_{X|\mathbf{c}^*}) \\
&= \frac{\sum_{x\in\Omega_X} P(\alpha|x,\mathbf{c})P(x|\mathbf{c})P(d_{X|\mathbf{c}^*}|x,\mathbf{c})}{\sum_{\alpha,x\in\Omega_X} P(\alpha|x,\mathbf{c})P(x|\mathbf{c})P(d_{X|\mathbf{c}^*}|x,\mathbf{c})} \\
&= \frac{\sum_x P(\alpha|x,\mathbf{c})P(x|\mathbf{c})}{\sum_{\alpha,x} P(\alpha|x,\mathbf{c})P(x|\mathbf{c})} \\
&= P(\alpha|\mathbf{c})\,.
\end{aligned}$$

By Tr. (2), $P^*(x|\mathbf{c}^*) = P^{\lambda}_{X|\mathbf{c}^*}(x|\mathbf{c}^*)$ is obtained by replacing $P(d_X|x, \mathbf{c}^*)$ with $\lambda_{x|\mathbf{c}^*}$ in the proof of i) Success, for each $(x, \mathbf{c}^*)$. Proofs for conditions ii) to iv) may be obtained analogously, replacing $P_{X|\mathbf{c}^*}$ by $P^{\lambda}_{X|\mathbf{c}^*}$. □

Corollary 2. *By [27, Th.5], consider any target event α and let $P^*(\alpha)$ be the PMF resulting from Pearl-Adams conditioning on uncertain evidence on random variable X, given $\mathbf{C} = \mathbf{c}^*$. It holds $P^*(\alpha) = (P \circ_A P_{X|\mathbf{c}^*})(\alpha)$, for every event α.*[6]

Example 14 (Ex. 7 Continued)**.** *Consider again the example of the Professor Disease (see p.35). The doctor decides to augment the BN representing her system of knowledge with auxiliary binary leaf node $D_{C|a}$. Based on her prior beliefs, $P(C|a) = \{0.60, 0.40\}$. The CPT of $D_{C|a}$ is specified following Def. 28:*

$$\begin{cases} P(D_{C|a} = d_{C|a}|a, c) = \eta \cdot 0.75/0.60 \\ P(D_{C|a} = d_{C|a}|a, \neg c) = \eta \cdot 0.25/0.40 \\ P(D_{C|a} = d_{C|a}|\neg a, c) = \eta \\ P(D_{C|a} = d_{C|a}|\neg a, \neg c) = \eta \end{cases}$$

It follows,

$$\begin{aligned} P^*(e) &= P(e|d_{C|a}) \\ &= \frac{P(e, d_{C|a})}{P(d_{C|a})} \\ &= \frac{\sum_{A,C} P(A)P(C|A)P(e|C)P(d_{C|a}|A, C)}{\sum_{A,C} P(A)P(C|A)P(d_{C|a}|A, C)} \\ &= \frac{0.342\frac{0.75}{0.6} + 0.141\frac{0.25}{0.4} + 0.001 + 0.021}{0.450\frac{0.75}{0.6} + 0.300 \cdot \frac{0.25}{0.4} + 0.002 + 0.247} \\ &= 0.538 \\ &= \left(P \circ_A P_{C|a}\right)(e) \end{aligned}$$

This result is the same we obtained previously by Adams conditioning.
Let $\alpha = (A = a)$, by Pearl-Adams conditioning we get $P^*(\alpha) = P(a|d_{C|a}) = 0.75 = P(\alpha)$. That is, conditional soft evidence on event $(c|a)$ does not change prior knowledge on conditioning event $(A = a)$; the postulate of Rigidity of contexts *is satisfied. Fulfillment of all kinematical postulates from Def. 17 may be checked analogously by simple calculations.*

As shown by Example 14, and by Th. 8, Pearl-Adams conditioning extends the equivalence result of [35] to the conditional setting. That is, Pearl's method is generalized by Pearl-Adams conditioning, just like Jeffrey's rule is generalized by

[6] If uncertain evidence on X is virtual, Adams conditioning is on $P^{\lambda}_{X|\mathbf{c}^*}$ from Tr. (2).

Adams conditioning. We stress absorption of marginal uncertain evidence, comes naturally as an instance of the conditional case, when $\mathbf{C} = \emptyset$. Also, full conditional uncertain evidence is nothing but conditional instances on every element of $\Omega_{\mathbf{C}}$.

3.2.2 Generalized Pearl's Method with Credal Uncertain Evidence

Credal virtual evidence (CVE) is obtained by replacing standard VE with intervals. Notation Λ_X is used here for the intervals $\{\underline{\lambda}_x, \overline{\lambda}_x : x \in \Omega_X\}$. To give an intuition, standard VE may be used to model partially reliable sensors or tests, whose quantification is based on sensitivity and specificity data. Indeed, these data are not always promptly available; e.g. a pregnancy test whose failure can be only decided later. When few data are available, a CVE with interval likelihoods can be quantified by the *imprecise Dirichlet model* (see [23] for details). Given N observations of r.v. X, if $n(x)$ of them reports x, the lower and upper bound of $P(x)$ for the imprecise Dirichlet model are, respectively:

$$\frac{n(x)}{N+s}, \quad \frac{n(x)+s}{N+s}$$

with s effective prior sample size. Consider the following example:

Example 15 ([171]). *The reference standard for diagnosis of anterior cruciate legament sprains is arthroscopy is called Declan test [40]. In a trial, 40 patients coming in with acute knee pain are examined using this test. Every patient also has an arthroscopy procedure for a definitive diagnosis. Results are hereby reported:*

	Declan Positive	*Declan Negative*
Arthroscopy Positive	*17*	*6*
Arthroscopy Negative	*3*	*14*

Indeed, 17 *cases are true positive,* 6 *are false negative,* 3 *are false positive and* 14 *are true negative. Based on available data, Declan test has 73.9% sensitivity rate (true positive cases over all positive cases), and 82.3% specificity (true negative cases over all negative cases), with overall 77.7% accuracy (percentage of rightly diagnosed cases). Any given patient visiting a clinic is given sprain probability* $P(x) = 0.2$. *Given a positive Declan test result, the imprecise Dirichlet model with* $s = 1$ *corresponds to a CVE with:*

$$\underline{\lambda}_x = \frac{17}{24}, \quad \overline{\lambda}_x = \frac{18}{24}, \quad \underline{\lambda}_{\neg x} = \frac{3}{18}, \quad \overline{\lambda}_{\neg x} = \frac{4}{18}.$$

The bounds of the revised sprain probability with respect to the above constraints are $\underline{P}^*(x) = \frac{1}{3}$, $\overline{P}^*(x) \simeq 0.53$. *A VE with frequentist estimates would have produced instead* $P^* \simeq 0.51$.

Transformation 3. *Given a BN over* $\mathbf{V}$ *and a CVE* Λ_X, *add binary child* D_X *of* X *and quantify its CCPT* $K(D_X|X)$ *with constraints* $\underline{\lambda}_x \leq P(d_X|x) \leq \overline{\lambda}_x$.

If X is binary, single constraint $l \leq P(x) \leq u$ defines a CS $K(X)$ with elements $P_1(X) := [l, 1-l]$ and $P_2(X) := [u, 1-u]$.
Tr. (3) reduces CVE updating in a BN to CN updating; see Fig. 3.1(left panel) as an example. The following holds:

Theorem 9. *Given a CVE in a BN, consider the CN returned by Tr.* (3). *For any event* α *it holds:*

$$\underline{P}(\alpha|d_X) = \underline{P}_{\Lambda_X}(\alpha)$$

where $\underline{P}_{\Lambda_X}(\alpha) = (P \circ_P \underline{\lambda}_x)(\alpha)$. *Analogous reasoning yields the upper bounds.*

Proof. The result follows from the analogous result with BNs. Without loss of generality, let $\alpha = \{X_Q = x_Q\}$, for some $X_Q \in \mathbf{V}$. For any BN consistent with the CN returned by Tr. (3), the conditional independence between X_Q and D_X given X implies:

$$P(x_Q|d_X) = \frac{\sum_x P(x_Q|x)P(x)P(d_X|x)}{\sum_x P(x)P(d_X|x)}\,. \tag{3.8}$$

The terms with set-valued specification in the right-hand side of Eq. (3.8) above are $\{P(d_X|x) : x \in \Omega_X\}$. The lower probability $\underline{P}(x_Q|d_X)$ according to the CN is therefore:

$$\min_{\substack{\lambda_x \in [\underline{\lambda}_x, \overline{\lambda}_x] \\ x \in \Omega_X}} \frac{\sum_x P(x_Q|x)P(x)\lambda_x}{\sum_x P(x)\lambda_x} \tag{3.9}$$

where the bounds of the optimization variables are those specified by Tr. (3). Eq. (3.9) is nothing but $\underline{P}_{\Lambda_X}(x_Q)$. This ends the proof. $\square$

Definition 29 (Pearl's Credal Method)**.** *Let* Λ_X *as above, and let* P *be a sharp PMF over* $\mathbf{V}$. *We define Pearl's Credal Method* $\circ_{CP}$ *as follows:*

$$(P \circ_{CP} \Lambda_X)(\alpha) = \left\{ P^*(\alpha) : P^*(\alpha) = \frac{\sum_{x \in \Omega_X} P(\alpha, x)\lambda_x}{\sum_{x \in \Omega_X} P(x)\lambda_x}, \lambda_x \in \Lambda_X, \forall x \in \Omega_X \right\}$$

Following previous reasoning,

Transformation 4. *Convert* Λ_X *into CSE* $K_X^{\Lambda}(X)$ *by the following:*

$$\underline{P}^{\Lambda}(x) = \frac{P(x)\underline{\lambda}_x}{P(x)\underline{\lambda}_x + \sum_{x' \neq x} P(x')\overline{\lambda}_{x'}} \tag{3.10}$$

and analogously, with a swap between lower and upper likelihoods, for the upper bound. Conversely, CSE $K'_X(X)$ is transformed into CVE Λ_X^K as follows:

$$\underline{\lambda}'_x \propto \frac{\underline{P}'(x)}{P(x)} \tag{3.11}$$

where $\underline{P}'(x) = \min_{P'_X(X)\in K'_X(X)} P'_X(x)$, and analogously for the upper bound.

Fig. 3.1(left panel) represents augmentation of a BN by credal nodes when set-valued marginal uncertain evidence is propagated by credal generalizations of Pearl's Method. We argue extension of the results from this section to the case of conditional CSE (right panel of Fig. 3.1) is straightforward, also based on Sec. 3.2.1.
We define the *shadow* of $K(X)$ the CS $\hat{K}(X)$ obtained from all the PMFs $\hat{P}(X)$ such that, for each $x \in \Omega_X$:

$$\min_{P(X)\in K(X)} P(x) \leq \hat{P}(x) \leq \max_{P(X)\in K(X)} P(x) \,.$$

A CS coinciding with its shadow is called *shady*. It is a trivial exercise to check that CSs over binary variables are shady. The following result provides the credal generalization of Prop. 2.

Theorem 10. *Absorption of a CSE with shady CSE $K'_X(X)$ is equivalent to that of CVE $\Lambda^K_{X_n}$, and,* vice versa, *absorption of a CVE Λ_X is equivalent to that of a CSE $K^\Lambda_X(X)$, as from Tr.* (4)*.*

Proof. To prove the first part of the theorem consider a VE λ_X consistent with the bounds in Eq. (3.11) and its analogous for the upper bounds, i.e. for each $x \in \Omega_X$:

$$\eta \cdot \frac{\underline{P}'(x)}{P(x)} \leq \lambda_x \leq \eta \cdot \frac{\overline{P}'(x)}{P(x)} \tag{3.12}$$

where η is the constant of proportionality making all the likelihoods smaller than one. By a derivation similar to that in the proof of Th. 9, we can express the CVE absorption as the optimization in Eq. (3.9) with the constraints in Eq. (3.12). After substitution $P'(x) := \eta^{-1}P(x)\lambda_x$ for each $x \in \Omega_X$, we obtain:

$$\underline{P}(x_Q|d_X) = \min_{\substack{\underline{P}'(x)\leq P'(x)\leq \overline{P}'(x)\\ x\in\Omega_X}} \sum_{x\in\Omega_X} P(x_Q|x)\cdot P'(x) \tag{3.13}$$

which corresponds to the absorption of the shadow of $K'(X)$. As the CS is shady, this proves the first part of the theorem. To prove the second part of the theorem, as a consequence of Pr. 2, each VE consistent with the CVE can be converted in a SE by Tr. (1). The CS implementing the CSE equivalent to the CVE is therefore:

$$K^\Lambda_X(X) := \left\{ P'(X) : P'(x) = \frac{P(x)\lambda_x}{\sum_x P(x)\lambda_x}, \underline{\lambda}_x \leq \lambda_x \leq \overline{\lambda}_x, \forall x \in \Omega_X \right\} .$$

Computation of $\underline{P}'(x)$ is a linearly constrained linear fractional task. If $P(x) > 0$, we can rewrite the objective function as:

$$P'(x) = \left[1 + \sum_{x' \neq x} \frac{\lambda_{x'} P(x')}{\lambda_x P(x)}\right]^{-1} \tag{3.14}$$

As $f(\gamma) = (1+\gamma)^{-1}$ is a monotone decreasing function of γ, minimizing the objective function in Eq. (3.14) is equivalent to maximize:

$$\sum_{x' \neq x} \frac{\lambda_{x'} P(x')}{\lambda_x P(x)} \tag{3.15}$$

and *vice versa* for the maximization. As each λ_x can vary in its interval independently of the others, the maximum of the function in Eq. (3.15) is obtained by maximizing the numerator and minimizing the denominator, i.e. for $\lambda_{x'} = \overline{\lambda}_{x'}$ and $\lambda_x = \underline{\lambda}_x$. This proves Eq. (3.10), which remains valid also for $P(x) = 0$. □

By Th. 9 and 10 credal belief revision of a BN is reduced to standard updating in a CN. For CSEs with non-shady CSs, the procedure is slightly more involved, as detailed by the following result, whose proof is analogous to that of the first part of Th. 10:

Proposition 3. *Given a CSE $K'_X(X) = \{P'_i(X) : i = 1, \ldots, k\}$ in a BN, add binary leaf node D_X of X, quantified by an ECPT $\{P_i(D_X|X) : i = 1, \ldots, k\}$ such that $P_i(d_X|x) \propto \frac{P'_i(x)}{P(x)}$ for each $i = 1, \ldots, k$ and $x \in \Omega_X$. Then:*

$$\underline{P}^*(\alpha) = \underline{P}(\alpha|d_X) \tag{3.16}$$

for any event α.

To clarify these results, consider the following example.

Example 16 (Ex. 13 Continued)**.** *The original PMF $P(X)$ is revised by a CSE with $K'_X(X) = \{P'_1(X), P'_2(X)\}$, where $P'_1(X) = [0.6, 0, 0.4]$ and $P'_2(X) = [0.4, 0, 0.6]$. $K'_X(X)$ is clearly a shady CS. Following Th. 10, this yields CVE $\Lambda_X^K = \{$2-3 : 1 : 8-12$\}$. Vice versa, a slightly different CVE $\tilde{\Lambda}_X := \{3-5 : 1 : 8-10\}$ induces the revised values $\underline{P}^*(g) = \frac{3}{5}$, $\overline{P}^*(g) = \frac{2}{3}$, $\underline{P}^*(y) = \underline{P}^*(y) = 0$, and $\underline{P}^*(r) = \frac{1}{3}$, $\overline{P}^*(r) = \frac{2}{5}$. These bounds can be equivalently obtained in a two-node CN with d_X child of X and CCPT $K(d_X|X)$ such that $P(d_X|X = g) \in [0.6, 1]$, $P(d_X|X = y) = 1$, and $P(d_X|X = r) \in [0.8, 1]$. Alternatively, following Pr. 3, the absorption of $K'(X)$ can be achieved by specifying an ECCPT made of two CPTs.*

Both conservative updating and CIR were introduced in Sec. 2.3. Particularly, CIR refers to the case of incomplete data, and account for missing data as either MAR (or CAR), or due to some mechanism on which *near-ignorance* is expressed [78].
Remember $\mathbf{X}_Q$ denotes a collection of *target* random variables, with $\mathbf{x}_Q \in \Omega_{\mathbf{X}_Q}$ target event. Also, $\mathbf{E}$ is the subset of fully observed nodes of the network, whereas nodes in $\mathbf{M} = \mathbf{M}_r \cup \mathbf{M}_{nr}$, with $\mathbf{M}_r \cap \mathbf{M}_{nr} = \emptyset$, were not observed as they are missing at random ($\mathbf{M}_r$) or due to some nearly-ignored mechanism ($\mathbf{M}_{nr}$).
Given a BN, CIR may be implemented in a CN obtained by adding $|\mathbf{M}_{nr}|$ auxiliary leaf nodes to the DAG, such that each variable in $X \in \mathbf{X}_{nr}$ is the unique parent of binary credal node X', with $\Omega_{X'} = \{x', \neg x'\}$ [10]. Let $\mathbf{X}'$, defined on $\Omega_{\mathbf{X}'}$, denote the collection of all auxiliary leaf (vacuous) nodes, transforming the BN into a CN; $\mathbf{x}' = \{X' = x' : X' \in \mathbf{X}'\}$. The ECCPT of each node $X' \in \mathbf{X}'$ is specified as:

$$\left\{ \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 1 & 1 \end{bmatrix}, \dots, \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \right\}.$$

Given $\mathbf{X}_Q = \mathbf{x}_Q, \mathbf{E} = \mathbf{e}$ and $\mathbf{M}$, Eq. 2.5 is reproduced by $K(\mathbf{x}_Q|\mathbf{e}, \mathbf{x}'_{nr})$, the CS obtained by instantiation of credal nodes $\mathbf{X}'$ and precise nodes $\mathbf{E}$ of the network. If the network is already credal, the CCM transformation may be applied to make it sharp; then CIR may be implemented with respect to every BN; see [10] for details on the procedure with locally and globally specified CNs. As a straightforward application of Th. 10, the CVE implementing CUR corresponds to a CSE associated to a vacuous CS. In fact, for $\underline{\lambda}_x = 0$ and $\overline{\lambda}_x = 1$, for each $x \in \Omega_X$, then Eq. (3.10) gives zero for the lower bound of $P'(x_n)$ and we similarly reach one for the upper bound. Credal Jeffrey's Rule becomes the CUR rule of [75] that represents the most conservative approach to belief revision:

$$\underline{P}'(\alpha) = \min_{x \in \Omega_X} P(\alpha|x) \tag{3.17}$$

We can similarly proceed in the case of *incomplete* observations, i.e. some values of X_n are recognized as impossible, but no information can be provided about the other ones. If this is the case, we just replace Ω_X with $\Omega'_X \subset \Omega_X$, as in Eq. (2.5).

The following proves consistency of Credal Pearl's Method for shady CSE:

Theorem 11. *Given a BN over $\mathbf{V}$ and a shady CSE $K'_X(X)$, convert the CSE into a CVE as in Tr. (4) and transform the BN into a CN by Tr. (4). Let $K(\mathbf{V}, D_X)$ be the joint CS associated to the CN. Then, $K(\mathbf{V}|d_X)$ comes from $P(\mathbf{V})$ by CPK on the partition induced by X. Moreover $K(X|d_X)$ coincides with the marginal CS computed in the CN.*

Proof. The result easily follows from the analogous result for PK, which holds for each $P'_X(X) \in K'_X(X)$. The consistency between $K(X|d_X)$ in the CN and $K'_X(X)$ in the original CSE specification follows from Th. 9. $\square$

Let us first stress how Tr. 4 does not affect the topology (nor the treewidth) of the original $\mathcal{G}$. As standard BN updating of polytrees can be performed efficiently, the same happens with uncertain instances. Similarly, with multiply connected models, standard BN updating is exponential in the treewidth, and nothing changes if the model is augmented by virtual(ized) evidence. As already discussed in Sec. 1.3.2, ApproxLP tackles CN updating based on linear programming. The algorithm reduces CN updating to a sequence of linear programming tasks, each obtained by iteratively fixing all the local models to single elements of the corresponding CSs, while leaving one single variable free. This provides an inner approximation of the updated intervals with the same complexity of a BN inference on the same graph. Remarkably, whenever a CN has all local CSs made of a single element apart from one, ApproxLP produces exact inferences. This is just our case, when coping with a single CVE or CSE, and ApproxLP might be therefore used to efficiently revise beliefs. Finally, it is obvious that CSEs on a root node can be trivially elicited by replacing the original, unconditional, PMF of the BN, with the CS associated of the CSE.

In the next section, we move a step forward and consider uncertain belief propagation with a CN. Up to this point, we implemented revision rules from Ch. 2 with BNs. When uncertain evidence is specified by a credal set (as it is the case with CSE), the network resulting from the propagation process corresponds to a special kind of CN, with a single credal node, and all others degenerate to single CPTs.

3.3 Uncertain Belief Propagation with Credal Networks

Consider the case of a single virtual instance - a CVE - on r.v. $X \in \mathbf{V}$. Its embedding in a CN is trivial and it just requires augmenting the latter with an auxiliary boolean child. Remarkably, absorption of a CVE by a CN enjoys two main properties for virtual findings in general:

1. The procedure naturally extends to the case of several CVEs - a dummy child is introduced for each node involved by the revision process[7];

[7] We will return to this in Ch. 5

2. Augmentation of the CN by such leaf nodes does not affect the topology of the network, e.g. if the graph is singly connected, complexity of inference is unchanged.

Embedding a CSE in a CN requires a more detailed discussion. Let $K(\mathbf{V})$ be the strong extension of the CN, the generalization to CNs of Eq. (3.13) becomes:

$$\underline{P}'_X(x_Q) = \min_{P(\mathbf{V})\in K(\mathbf{V})} \min_{P'(X)\in K'(X)} \sum_x P(x_Q|x)P'(x)\,, \tag{3.18}$$

that is, to Eq. (2.7). Analogous reasoning applies to the upper envelope $\overline{P}'_X(x_Q)$, where maximization is considered instead.
If CSE is gathered on a root node, the multilinear task of Eq. (3.18) above trivially consists in replacing the existing CCPT - $K(X)$ - by the former - $K'_X(X)$. In the general case, straightforward extension of previous proposals would yield augmentation of the DAG by auxiliary credal node D_X, whose CCPT is specified by:

$$\underline{P}(d_X|x) \propto \underline{\lambda}_X(x) = \frac{\underline{P}'(x)}{\overline{P}(x)}\,. \tag{3.19}$$

There, $\underline{P}'(x)$ and $\overline{P}(x)$ are obtained as the lower envelope of $K'_X(X)$, and as the upper envelope of $K(X)$ at x, respectively, for each $x \in \Omega_X$.
It is easy to see instantiation of node D_X to its *truth* value d_X yields a (likely very loose!) CS $K''_X(X) \supseteq K'_X(X)$. We propose the following to convert a CSE into a CVE for a given CN:

Transformation 5. *For a fixed $x \in \Omega_X$, convert CSE into the CVE Λ_X, with generic element:*

$$\underline{\lambda}_{i,j,x} = \frac{P'_{i,X}(x)}{P_j(x)}$$

for all $i = 1, \ldots, |extK'_X|$, and $j = 1, \ldots, |extK|$, for each $x \in \Omega_X$.

Consider Fig. 3.2, and suppose our event of interest α corresponds to target state x_Q of the single query r.v. $X_Q \in \mathbf{V}$[8]. Belief propagation by the generalized Pearl's method on a CN is likely intractable, as it requires enumeration of all vertices minimization of the following non-linear problem:

$$\underline{P}(x_Q|d_X) = \min_{\substack{P_j(\mathbf{V})\in K(\mathbf{V}) \\ P_i(X)\in K'_X(X)}} \frac{\sum_{x\in\Omega_X} P_j(x_Q,x)\lambda_{i,j,x}}{\sum_{x\in\Omega_X} P_j(x)\lambda_{i,j,x}}\,, \tag{3.20}$$

[8]Considering a single query r.v. will allow derivation of desirable complexity results, see for example Prop. 8 of Ch. 5.

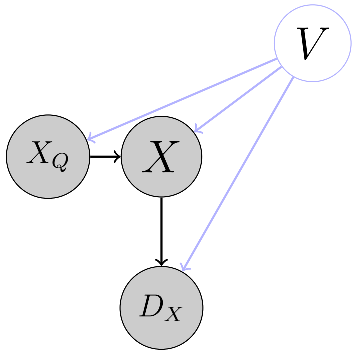


Figure 3.2: Equivalent representation of the augmented CN by auxiliary binary node D_X, as a collection of underlying BNs (grey), indexed by transparent node V.

where $P_j(x_Q, x)$ requires in turn enumeration of all extreme points of those CSs involved by the minimization task, namely $K(X_Q)$ and $K(X|X_Q)$ (or $K(X|x_Q)$, if the CCPT is separately specified). Theoretically, this task shall be tackled by introducing a so-called *transparent* node V. This is uniformly distributed, and simultaneously indexes all extreme points of each CCPT of the network; see the toy CN of Fig. 3.2 as an example. Introduction of such a node allows capturing all imprecision from the model. As a consequence, the underlying DAG (in grey, in Fig. 3.2), would reduce to a BN. It is straightforward to see Eq. (3.20) may thus be equivalently written as:

$$\underline{P}(x_Q|d_X) = \min_{v\in\Omega_V} \frac{\sum_{x\in\Omega_X} P(x_Q, x|v)P(d_X|x, v)}{\sum_{x\in\Omega_X} P(x|v)P(d_X|x, v)}\,.$$

Suppose a CN has $(n+1)$ nodes, each associated with a k-variate r.v., introduction of auxiliary node D_X and of transparent node V requires $|\Omega_V| = k^{(n+2)}$; in Fig. 3.2, $|\Omega_V| = k^3$.

We argue representation of the propagation setup as that from Fig. 3.2 provides an intuitive justification of the fulfillment of the conservativeness postulate of CPK (Def. 15) by the Generalized Pearl's Method for CNs. Nevertheless, inference is likely intractable and we shall resort to an approximate approach.

An outer approximation to the solution to such a multilinear problem, can be achieved by separately solving the two minimization problems, with respect to i) $K(\mathbf{V})$ and ii) $K'_X(X)$.
Consider minimizing the conditional terms first. It holds:

$$\underline{P}'_X(x_Q) \leq \min_{P'(X)\in K'(X)} \sum_x \underline{P}(x_Q|x)P'(x)\,. \tag{3.21}$$

In other words we can approximate the CVE computation by updating the original credal network for each possible value, and then solving the linear programming

task above. We denote the solution to Eq. (3.21) with $\underline{P}_1(x_Q)$.
A second solution is obtained by minimizing the CSE first, then the conditional terms. This case shall also be considered by rewriting:

$$\begin{aligned}\underline{P}'_X(x_Q) &= \min_{P(\mathbf{V})\in K(\mathbf{V})} \min_{P'(X)\in K'(X)} \sum_x \frac{P(x_Q,x)P'(x)}{\sum_{x_Q} P(x_Q,x)}\\ &= \min_{P(\mathbf{V})\in K(\mathbf{V})} \min_{P'(X)\in K'(X)} \sum_x P'(x) \frac{1}{1+\sum_{x'_Q\neq x_Q} P(x'_Q,x)}\\ &\geq \min_{P(\mathbf{V})\in K(\mathbf{V})} \sum_x P'(x) \min_{P'(X)\in K'(X)} \frac{1}{1+\sum_{x'_Q\neq x_Q} P(x'_Q,x)}\\ &= \min_{P(\mathbf{V})\in K(\mathbf{V})} \sum_x P'(x) \min_{P'(X)\in K'(X)} \frac{1}{2-P(x_Q,x)}\\ &= \min_{P'(X)\in K'(X)} \sum_x P'(x) \frac{1}{2-\overline{P}(x_Q,x)}\,,\end{aligned}$$

where the first line of equations is just Eq. (3.21). We denote last line of equations with $\underline{P}_2(x_Q)$.
Overall, we run both approximation and take the one giving the tightest approximation, that is:

$$\underline{P}^*(x_Q) = \max\{P_1(x_Q), P_2(x_Q)\}\,.$$

This represents an approximate approach to the updating of credal soft evidence in a CN. Nonetheless, as a drawback, the object associated with the approximate solution may no longer be accounted for as a *static* CN, unless X is a root node. Extension of the multilinear technique of [58] to the case of credal uncertain evidence might prove as an alternative viable approach to propagation of uncertain beliefs in CNs. Also, a naïve alternative worth mentioning is taking the CVEs obtained by Tr. 5:

$$\underline{\lambda}^*_x = \frac{\underline{P}'_X(x)}{\underline{P}(x)}, \quad \overline{\lambda}^*_x = \frac{\overline{P}'_X(x)}{\overline{P}(x)}\,.$$

On the one hand, usage of such loose bounds for the CVE would reduce complexity of the optimization task, and produce a static CN. On the other hand, the imprecision they yield would likely have a detrimental impact on the overall procedure.

This chapter provided several tools for belief propagation in DAG-based models, when uncertain evidence is provided. Our proposals extend Pearl's method for VE, and were proved to be sound under a kinematical point of view. Future work will jointly consider other non-kinematical approaches for comparisons, such as those based on maximum entropy [197]. Also, our focus will be on the graphical implementation of Imaging operators proposed in Ch. 2.
As for this last section on CNs, future developments will tackle methods and applications of belief propagation techniques with CNs when uncertain evidence is provided.

Those will include, among others, extension of Da Rocha *et al.*'s approach to the case of CSE, as well as approximated heuristics.

Chapter 4

Belief Merging

Opinion pooling - or belief aggregation, or merging - extends AGM theory to the case of multiple agents, whose judgments are defined on shared domains. Pooling is performed by functionals, that combine elements from a collection of beliefs toward a shared consensus, or agreement. This topic is of general interest, e.g. in the literature of statistical and decision theory [116, 113].[1]

Formally, when a number of independent sources gather probabilistic evidence on the same domain, we move from the single to the multi-agent setting. Members of a pool are called interchangeably (*epistemic*) *peers*, *agents*, or *sources*.

Let $m \geq 1$ agents be part of a pool. If no reasons are known to choose one member's opinion over the others, provided that each is belief is consistent with the agent's own system, pooling operators (POs) seek to synthesize all contributions into a single belief. Such a task is not straightforward, as motivated by the following example, from [87].

Example 17. *A pool of* $m = 3$ *agents is asked to provide an aggregated opinion on the set of propositions* $\{\phi, \phi \rightarrow \psi, \psi\}$ *from some language* $\mathcal{L}$. *Each agent is equipped with a consistent belief over the set, as depicted below:*

	ϕ	$\phi \rightarrow \psi$	ψ
Ag.1	*1*	*1*	*1*
Ag.2	*1*	*0*	*0*
Ag.3	*0*	*1*	*0*

[1] Through the past decades we assisted to a massive production of contributions to the subject of *belief fusion*; see, e.g. [22, 162, 178]. This latter term is often used interchangeably with those above. For clarity, we shall (arbitrarily) refer to belief fusion as the general task of aggregating two or more knowledge bases, rather than that of merging opinions from a pool on a sub-domain of Your belief.

where 1 *and* 0 *denote (boolean) True and False, respectively. If simple majority voting was adopted as pooling criterion, the following would result:*

ϕ	$\phi \rightarrow \psi$	ψ
1	*1*	*0*

This is known as discursive dilemma in [87]. In words, while each opinion is itself consistent, the aggregated opinion might not be so in the general case.

In our probabilistic framework, each peer provides a judgment over the (finite) possibility space of some r.v. X. This way, source j produces a PMF P_j over Ω_X, such that, most likely, $P_j(x) \neq P_{j'}(x)$ for some $x \in \Omega_X$ and $j' \neq j$, provided that $\sum_{x \in \Omega_X} P_j(x) = 1$, for each j, $j = 1, \ldots, m$. Let $|\Omega_X| = k$, the pool is associated a *belief profile*, i.e. matrix $\mathbf{P}^m(X) \in \Delta^{m \times k}$:

$$\mathbf{P}^m(X) = \begin{bmatrix} P_1^T(X) \\ \cdots \\ P_j^T(X) \\ \cdots \\ P_m^T(X) \end{bmatrix} = \begin{bmatrix} P_1(x_1) & \ldots & P_1(x_k) \\ \cdots & \cdots & \cdots \\ P_j(x_1) & \ldots & P_j(x_k) \\ \cdots & \cdots & \cdots \\ P_m(x_1) & \ldots & P_m(x_k) \end{bmatrix} .$$

$\mathbf{P}^m$ is a *coherent* belief profile if there exists at least one element of Ω_X whose assessed probability differs from zero for all agents; it is *incoherent* otherwise [86].
Analogously to $\mathbf{P}^m$, $\mathbf{K}^m(X)$ is the matrix whose columns are CSs $K_1(X), \ldots, K_m(X)$. We use simplified notation $\mathbf{K}$ to denote $\mathbf{K}^m$ (and $\mathbf{P}$ for $\mathbf{P}^m$) when cardinality of pool is clear from the context. For a lower probability space, it is reasonable - to our purposes - to assume events are defined on a finite σ-algebra.[2] We refer to sharp and imprecise belief profiles to characterize the two different setups, even though it must be kept in mind that all techniques referring to the first shall be intended as degenerate cases of those set-valued.
As a remark, imprecise opinions may be produced by noisy observational processes upstream, unreliable communication, qualitative or semi-qualitative judgments, and so forth. Additional types of assessments may be gathered to the pool, including odds ratios, or likelihood functions; we will return to these in Sec. 4.2.1.
Usage of pooling techniques was motivated in the literature as pursuing either:

a) Symmetric Belief Merging: POs merge a collection of opinions held by doxastic (or rational) agents, with the aim to i) describe a collection of event, ii) perform collective decision making, and/or iii) improve shared knowledge among members of the pool, as if acting like a single agent;

[2]See [85] on infinite σ-fields, and [88] on general agendas, i.e. algebras closed under negation only.

b) Asymmetric Belief Merging: The PO updates a prior distribution into a posterior. In the decision making setup, a further expert is consulted upon conflicting beliefs, considered as evidence [185, 187]. Such a *supra-Bayesian* agent updates her belief Ω_X. Supra-Bayesian POs may also refer to a single member of the pool, apprised of the remaining $m-1$ opinions.

A selected review of POs is provided in Sec. 4.2. These are characterized as either fully sharp[3], sharp-to-imprecise, imprecise-to-sharp or fully imprecise. We define a general PO as the mapping:

$$\Pi : \mathcal{P}_X^m \to \mathcal{P}_X \, . \tag{4.1}$$

$\Pi_{\mathbf{K}}(x)$ (or just $\Pi(x)$, when clear from the context) denotes $\Pi(\mathbf{K})(x)$, pooling of profile $\mathbf{K}(X)$, evaluated at some $x \in \Omega_X$. The definition provided by Eq. (4.1) underlies two so-called *regularity* conditions for a PO, according to [87]:

- Universal Domain: Π is defined on $\mathcal{P}_X$, set of all possible coherent PMFs on r.v. X;

- Collective Rationality: Π generates coherent PMFs on r.v. X.

4.1 Principles for Opinion Pooling

This section provides an overview of desirable properties for POs. Quite ironically, no consensus was reached so far on a single pooling operator as *optimal* under all circumstances, nor on which principles to choose as essential. It was proved, among others by Dietrich and List [87], that no PO satisfies all possible desiderata under all circumstances. The authors accounted for four key principles for a PO: Universal Domain, Collective Rationality, Anonymity (Def. 32) and Unanimity (Def. 35, and proved the only functional that satisfies them all fails *Non-Dictatorship*. A PO fails such principle whenever a single opinion fully determines the aggregated belief, irrespective to changes in the others'. Trivially, if fulfillment of Non-Dictatorship was questioned, the whole pooling process would no longer be worth consideration. The result in [87] generalizes the best known Arrow's impossibility theorem that any constitution satisfying transitivity, independence of irrelevant alternatives, and unanimity is a dictatorship [13, 14]. Several related works proposed dropping one of the four principles, to guarantee soundness of the pooling procedure; e.g. [202] proposed Anonymity ought to be dismissed as a requirement.

[3]See also [238] for an up-to-date overview.

Before we proceed and list a number of desirable properties of a PO, let us point out (probabilistic) belief aggregation, or merging, may be intended as a generalized form of (probabilistic) belief revision. While the latter, based on the AGM postulates, seeks to study how a doxastic agent ought to adjust her deductively closed set of propositions upon newly acquired information, opinion pooling tackles combination of several beliefs toward a consensus. Remarkably, Konieczny and Pino Pérez [152] proposed six postulates, acting as the counterpart to KM1-KM6 from Ch. 2. In our framework, the pool expresses on a coarse partition of Ω, say Ω_X, rather than on the whole collection of events in Σ, and postulates translate as follows:

KP1 $\Pi_{\mathbf{K}}(x)$ is a CS,[4]

KP2 If $\cap_{j=1}^{m} K_j(X) \neq \emptyset$, then $\Pi_{\mathbf{K}}(X) = \cap_{j=1}^{m} K_j(X)$,

KP3 If each $K_{1,j} \equiv K_{2,j}$, $K_{i,j} \in \mathbf{K}_i$, for each $i = 1, 2$ and $j = 1, \dots, m$, then $\Pi_{\mathbf{K}_1}(X) \equiv \Pi_{\mathbf{K}_2}(X)$,

KP4 If $\left(\cap_{j=1}^{m} K_{1,j}(X)\right) \bigcap \left(\cap_{j=1}^{m} K_{2,j}(X)\right)$ is the empty set, then $\Pi_{\mathbf{K}}(X) \not\subseteq \Pi_{\mathbf{K}_1 \bigsqcup \mathbf{K}_2}(X)$, where $\mathbf{K}_1 \bigsqcup \mathbf{K}_2 = \{K_{1,1}(X), \dots, K_{1,m}(X), K_{2,1}(X), \dots, K_{2,m}(X)\}$,

KP5 $(\Pi_{\mathbf{K}_1}(X) \cap \Pi_{\mathbf{K}_2}(X)) \subseteq \Pi_{\mathbf{K}_1 \bigsqcup \mathbf{K}_2}$,

KP6 If $\Pi_{\mathbf{K}_1}(X) \cap \Pi_{\mathbf{K}_2}(X) \neq \emptyset$, then $\Pi_{\mathbf{K}_1 \bigsqcup \mathbf{K}_2} \supseteq (\Pi_{\mathbf{K}_1}(X) \cap \Pi_{\mathbf{K}_2}(X))$.

Let (Ω, Σ) be any measurable space, and Π be any PO taking a belief profile on r.v. X as input. As a first, a fully sharp PO is *linear* if it produces a convex combination of the elements of $\mathbf{P}$.

Definition 30 (Convexity). Π *satisfies convexity whenever every PMF consistent with the pool's opinion may be expressed as a convex combination of* $ext\Pi(X)$*. Formally,*

$$P(X) \in \Pi(X) \iff P(X) \in CH\{ext\Pi(X)\}\,.$$

When the pooling process involves sharp probabilities only, $\Pi(X)$ trivially satisfies convexity.[5] Consider the following example:

Example 18. *Suppose* $m = 3$ *sources provide a PMF on event* $(X = x)$ *resulting out of* 10 *Bernoulli trials, such that* $\Omega_X = \{x, \neg x\}$*:* $P_1(x) = 2/10, P_2(x) = 3/10$ *and* $P_3(x) = 4/10$.

[4] This corresponds to Dietrich and List's regularity principle of Collective Rationality.

[5] Also, any non-vacuous fully imprecise PO (Sec. 4.2.4) that satisfies convexity while generating a closed set of PMFs satisfies a further *Defining Principle* [2].

If the convex hull induced by the m opinions is considered as the aggregated pool's opinion, it holds $0.2 \leq P'(x) \leq 0.4$*,* $P'(X) \in \Pi(X)$*. Suppose X represents the tossing of a coin, and x =Heads:* 2.5 *heads out of* 10 *flip coins are evaluated as a consistent opinion from the pool. This type of information is likely accepted by a sensitivity analysis approach only, whereas others might refuse it.*

Some arguments in the spirit of Ex. 19 above may be posed against convexity as a feature of Π; indeed, we do not require it.

Definition 31 (Neutrality [253, 179])**.** *A PO is neutral if there exists some function g s.t.*

$$\Pi(X) = g(K_1(X), ..., K_m(X)) \, .$$

If events are defined on a σ-algebra, and every neutral fully sharp PO is linear [88], and vice versa [238].

Definition 32 (Anonymity [179, 88])**.** *If* $\mathbf{K}_1, \mathbf{K}_2$ *fully agree on X's behavior, while they differ on the whole domain induced by* $\mathbf{V}$*, Anonymity requires:*

$$\Pi_{\mathbf{K}_1}(X) = \Pi_{\mathbf{K}_2}(X) \, .$$

Anonymous POs are intended as opposed to a *holistic* understanding of the pool's dynamics [88]. Such principle prevents *Dutch Bookies* from influencing the collective opinion, e.g. by leveraging misleading arguments. When a PO is not influenced by the agents' belief base beyond X, it defines a mapping that is invariant to any ordering. Invariance with respect to permutations of the elements in the belief profile may be specified as a sub-principle, namely of *Equivalence* or *Atomic Permutation* (see [2] for details).

Definition 33 (Marginalization [238, 179, 116])**.** *When a belief merging is on joint r.v.* $\mathbf{X}$*, any PO commuting with marginalization satisfies this principle. Without loss of generality, let* $\mathbf{X} = \{X, Y\}$*, it holds:*

$$\Pi^{\downarrow Y}(\mathbf{X}) = \Pi({K_1}^{\downarrow Y}(\mathbf{X}), ..., {K_m}^{\downarrow Y}(\mathbf{X})) \, , \tag{4.2}$$

and analogously for $\Pi^{\downarrow X}(\mathbf{X})$*.*

An equivalent characterization was provided by [116]; see also [238] for a further definition. It was proved by [179] marginalization is equivalent to an extended form of Neutrality, requiring the existence of a function g', such that $\Pi(X) = g'(X, K_1(X), ..., K_m(X))$.

Definition 34 (Consistency [264])**.** *Any PO satisfying Consistency lies in the convex hull induced by the belief profile.*

A range of specializations for consistency, or *preservation of initially shared agreements*, may be found in the literature, including Indifference Preservation[6] [85], Zero Preservation[7] [179, 238], and Strong Consistency, this latter requiring reachability of the bounds. Linear pooling functions are the only satisfying Zero Preservation [253, Th.6.7] and Anonymity, provided $|\Sigma| \geq 3$ [179]. Among others, we define Unanimity based on [87]:[8]

Definition 35 (Unanimity)**.** *Let* $\mathbf{K}(X)$ *be a belief profile such that it holds* $P(X) \in K_j(X)$*, for each* $j = 1, \ldots, m$*. By Unanimity, we expect* $\Pi(X) \ni P(X)$*.*

It is straightforward to see Def. 35 extends KP2 above.

Definition 36 (Collegiality [2])**.** *Let some ordering be available on the elements of the belief profile, such that the (consistent) conjunction of all opinions of the last* $m - k$ *experts are consistent with the aggregated opinion of the first* k*'s. Formally,*

$$\Pi_{\mathbf{K}^{1,\ldots,k}}(X) \subseteq \cap_{i=k+1}^{m} K_i(X) \neq \emptyset \,,$$

with $\mathbf{K}^{1,\ldots,k}(X) = \{K_1(X), \ldots, K_k(X)\}$*.*

Collegiality shall be intended as a special case of KP5. When it is satisfied, uninformative additional opinions ought not to yield dilation [196], so long they are consistent with those already accounted for.
If both Collegiality and Consistency are satisfied, $\Pi(X)$ maps the whole belief profile into $\cap_{i=1}^{k} K_i(X)$, if this is not empty [264, Lemma 2.1].
A strong criticism may be raised against Collegiality: in a situation when all peers but one, say j, provide vacuous opinions, an almost Dictatorship of the j-th agent would result.

Definition 37 (Agreement [264, 2])**.** *Suppose a belief profile consists of two subgroups of agents, whose aggregated opinions are consistent:* $\Pi_{\mathbf{K}^{1,\ldots,k}}(X) \cap \Pi_{\mathbf{K}^{k+1,\ldots,m}}(X) \neq \emptyset$*, with* $\mathbf{K}^{k+1,\ldots,m} = \{K_{k+1}(X), \ldots, K_m(X)\}$*. By the Agreement principle it holds:*

$$\Pi(X) = \Pi_{\mathbf{K}^{1,\ldots,k}}(X) \cap \Pi_{\mathbf{K}^{k+1,\ldots,m}}(X) \,.$$

Agreement strengthens KP6 above. By [2, Th.4.1.1], if also Strong Consistency is satisfied, Collegiality follows.

[6] Indifference Preservation is implied by Anonymity as transformation of the uniform credence function under permutation that yields the same aggregate belief.

[7] It refers to the case when the whole pool is certain on a negated event. When $|\Sigma| \geq 3$, this is implied by Neutrality [179]. Also, Zero Preservation and Anonymity together imply Neutrality [179, Th.3.2].

[8] See [198, 111] for alternative definitions.

If the two aggregated opinions are not consistent, and $\cap_{j=1}^{m} K_j(X) \neq \emptyset$, by the Strong Disagreement principle, it holds:

$$\Pi(X) \cap \Pi_{\mathbf{K}^{1,\ldots,k}}(X) = \emptyset\,,$$

thus dropping the condition of consistency among individual opinions. This result is enhanced by KP2.

Definition 38 (Ignorance [264]). *If $k < m$ sources from the pool hold vacuous opinions on r.v. X, by Ignorance it holds:*

$$\Pi(X) = \Pi_{\mathbf{K}^{k+1,\ldots,m}}(X)\,.$$

The following is a key principle to our graphical approach to probabilistic belief revision. It requires experts acting as if they were a single doxastic agent undergoing a learning process: that is, pooling commutes with updating [198, 88], or, following [252], with probabilistic revision. Formally:

Definition 39 (External Bayesianity [166, 113]). *Let $L : \Omega \to \mathbb{R}$ be any likelihood function, shared by all peers, such that $0 < \sum_{\omega:\omega\sim X} L(\omega)P(\omega) < \infty$. Also, let the transformed belief profile $\mathbf{K}^{L;m}(X)$ be defined by $\{\frac{L(X)P_j(X)}{\sum_{\omega\in\Omega} L(\omega)K_j(\omega)} : P_j(X) \in K_j(X), j = 1, \ldots, m\}$. Any PO is Externally Bayesian if it holds:*

$$\Pi_{\mathbf{K}^{L;m}}(X) = CH\left\{\frac{L(X)P'(X)}{\sum_{\omega\in\Omega} L(\omega)P'(\omega)} : P'(X) \in \Pi(X)\right\}.$$

Def. 39 above is also known as Conditionalization on L-information [85], further decomposed into Conditionalization on Public Information - all agents experience learning - as opposed to Private information - only one peer does. In [85, 237], Individualwise Bayesianity is also considered, generalizing Conditionalization on Private Information back to External Bayesianity. The general principle of *Conditionalization on Information*, requires the pooling process to commute with one or more peers experiencing a learning episode. By [85, Th. 4] no PO exists that satisfies Individualwise Bayesianity, under Non-Dictatorship; see the footnote for further remarks.[9]

[9] Specialization of single (Private), partial or collective (Public) learning experiences is motivated by a number of reasons in the literature. As a first, some believe accounting for everyone's opinion is not necessarily the best strategy as it may lower the collective knowledge toward fully shared awareness. Conversely, a single member's update in knowledge may significantly improve the group's performance as a whole. Yet, imposing knowledge (or belief) of a single, or a proportion, of peers to the whole group is somehow questionable. Although we reckon situations where members of the group are equipped with more knowledge than others may be of some interest, this is not

Definition 40 (Probabilistic Independence Preservation [198, 1, 238]). *Consider the case of a sharp belief profile. If it holds $P_j(X|y) = P_j(X)$, for some event y, with $P_j(y) > 0$, for each $j = 1, \ldots, m$, we require $\Pi_{\mathbf{P}|c}(X) = \Pi_{\mathbf{P}}(X)$, where $\Pi_{\mathbf{P}|y}(X) := \Pi(\mathbf{P})(X|y)$.*

By [253], the only POs satisfying both Neutrality and Probabilistic Independence Preservation fail Non-Dictatorship.[10] Extensions to the imprecise setting are straightforward, although it must be remembered independence concepts can be no longer used interchangeably (cfr Sec. 1.2.2).
We introduce Locality as the counterpart to CoPK axioms for belief merging:

Definition 41 (Locality [264]). *If a pool expresses on conditional event $(X|Y = y)$, knowledge of r.v. X conditional on any $y' \in \Omega_Y \backslash \{y\}$ ought to be left unchanged by any (asymmetric) PO satisfying Locality.*

A number of arguments supporting Locality are found in the literature of epistemic entrenchment, see, e.g. [90]. Failure of Locality is connected to the properties of Independence Preservation of the PO. Since we are interested in performing opinion pooling with probabilistic graphical models, we argue POs failing Locality would produce counterintuitive dynamics with respect to the pattern of independences among a set of variables, and ought to be avoided. Graphical implementation of a number of POs below provides a straightforward justification of their fulfillment of Locality, as well as of External Bayesianity.
Among others, further principles worth mentioning are *Family Aggregation*, that

our case. We believe learning experiences should always be evaluated at the pool level, relying on the peers being actually "expert" and receptive. Reliability of the single agent is a critical issue that should be addressed, rather than an approach based on single agents' wealth of experience. To this purpose, peer-specific weights might be used, possibly failing Anonymity. In this direction, we introduce Informed LogOp, and its graphical counterpart, in Sec. 4.2.1. Additionally, from a methodological perspective, any pooling operator that satisfies Conditionalization on non-Public Information is something that should be rejected upstream: if some knowledge is not shared, why should we expect it to be common ground after pooling?
As a further remark, consider the issue of the perception that each member has of the same learning experience. A full mathematical solution would require, as a starting point, a much wider range of possible events, making the problem infeasible, and likely gibberish. When it comes to updating, conditioning based on likelihood functions serves as a viable expedient: when an observation is affected, for many possible reasons, with some degrees of uncertainty, a cautious approach would specify the first by means of likelihood ratios [252] (or Bayes Factors, as with Wagner's rule).
All things considered, we argue straightforward allowance of single peers to act under non-uniform awareness of the system opens the way to dutch-bookies.

[10] It was proved that every non-trivial PO required to satisfy Non-Dictatorship and Anonymity, fails the Probabilistic Independence Preservation principle when $|\Omega| \geq 5$ [114]. A hierarchy of four weaker additional principles was proposed by [198].

applies to parents' sets in DAG-based graphical models[11], and *Proportionality*, requiring invariance of the pooled belief to replacement of each element from the profile by $m' > 1$ clones. Finally, fulfillment of the *Continuity* principle applies to sequences of $m \to \infty$ beliefs, converging to some $K^*(X)$. The interested reader may refer to [198, 264, 85], respectively, for details.

Next section introduces POs and their properties. Contributions from the section were previously published in [171].

4.2 Generalized Opinion Pooling Operators

As already outlined, we categorize POs in this section as either fully sharp (Sec. 4.2.1), sharp-to-imprecise (Sec. 4.2.2), imprecise-to-sharp (Sec. 4.2.3) and fully imprecise (Sec. 4.2.4), for clarity of exposition.

4.2.1 Fully Sharp Pooling Operators

A fully sharp PO maps any belief profile $\mathbf{P}(X) \in \Delta^{m \times k}$ into a single PMF.

Definition 42 (Linear (Pooling) Operator [239])**.** *Given an m-dimensional belief profile $\mathbf{P}(X)$, and a convex collection of non-negative weights w_j, the Linear PO (LinOP) is defined as:*

$$LinOp_{\mathbf{P}}(x) = \sum_{j=1}^{m} w_j P_j(x) \tag{4.3}$$

for each $x \in \Omega_X$. When $w_j = m^{-1}$, $j = 1, \ldots, m$, Eq. (4.3) *is called Arithmetic LinOp.*

LinOp is the only fully sharp PO that satisfies Anonymity [253, Th. 6.7], provided $|\Sigma| \geq 3$. Also, it satisfies Marginalization [238, Prop. 2], Consistency and Strong Monotonicity [111], while it fails Locality. This last feature will be relevant in Sec. 4.2.4.

Definition 43 (LogOp [17, 115])**.** *Given an m-dimensional belief profile $\mathbf{P}(X)$,*

$$LogOp_{\mathbf{P}}(x) = \frac{\prod_{i=1}^{m} P_i(x)^{\alpha_i}}{\sum_{x \in \Omega_X} \prod_{i=1}^{m} P_i(x)^{\alpha_i}} \tag{4.4}$$

for every $x \in \Omega_X$. Each weight α_j must satisfy $\alpha_j > 0$, for all $j = 1, \ldots, m$, and $\sum_{j=1}^{m} \alpha_j = 1$.

[11] By [198, Prop. 1], simultaneous fulfillment of both Family Aggregation and Consistency require failure of Non-Dictatorship.

Geometric pooling is widely used in practical settings (e.g. [41, 219]). LogOp requires all belief profiles to be coherent [88], to comply with KP1. Analogously to Π for $\Pi_{\mathbf{K}}$ (or $\Pi_{\mathbf{P}}$), we write LogOp for $\text{LogOp}_{\mathbf{P}}$, to avoid cumbersome notation; the same applies to all other POs.
LogOp always satisfies Strong Consistency [115, 85, 238], External Bayesianity [115, 16] (see also [102]), Indifference Preservation and Continuity [85, Th.1], and a weaker form of Independence Preservation, called Markov Independence Preservation [198]. Also, Ordinary LogOp - i.e. the special case of general LogOp, with all equal weights $\alpha_j = m^{-1}$ - simultaneously satisfies External Bayesianity, Indifference Preservation and Continuity [88]. Finally, when Dietrich's Conditionalization is on Public Information, the only fully sharp POs satisfying it is LogOp with $\alpha_j > 0$ for at least one j [85, Th.2] (and [216, Central Theorem]); this is always true according to our characterization of LogOp from Def. 43.
A further fully sharp PO, called MultOp was proposed in [88], as a variation of Eq. (4.4) with $\alpha_j = 1, j = 1, \ldots, n$. It satisfies Conditionalization on Private Information (and Indifference Preservation and Continuity, [85, Th.3]). Degenerate LinOp - with $w_j = 1$ for some j, and $w_{j'} = 0$ for every $j' \neq j$ - and MultOp fail Non-Dictatorship [85].

It is easy to see Ordinary LogOp yields the (normalized) geometric mean, while Arithmetic LinOp the arithmetic mean of a given vector $(a_1, \ldots, a_m)'$, where $a_j = P_j(x)$, for some $x \in \Omega_X$. Let $m = 2$, the arithmetic mean geometrically corresponds to the center of mass of the vector, while the geometric mean is given a more involved interpretation: it corresponds to the length of the tangent line shared by two circles, one of diameters a_1, a_2 the other, that are tangent externally. A thorough discussion on the geometric interpretation of the two POs is out of the scope of this work; what is relevant to our purposes is that both operators map the belief profile within the convex hull induced by $\mathbf{P}$. While LinOp yields the KL-projection of the pool's belief profile, i.e. the point of $\mathcal{P}_X$ that minimizes the KL divergence (Eq. (2.3), and the L_2 norm) among all members of the pool, LogOp solves the analogous minimization problem for the reversed KL-divergence [111]. Formally, it holds:

$$\begin{aligned} LinOp(x) &= \text{argmin}_{P(x)} \textstyle\sum_{j=1}^{m} KL(P_j(x) \| P(x))\,, \\ LogOp(x) &= \text{argmin}_{P(x)} \textstyle\sum_{j=1}^{m} KL(P(x) \| P_j(x))\,. \end{aligned} \tag{4.5}$$

As a further property, both POs guarantee that, if an additional expert joined the pool, *ceteris paribus*, the aggregated opinion would change in this direction, provided informativeness of her belief. This Monotonicity property [111] applies in its strong version to LinOp, while LogOp only satisfies a weaker form.

Graphical Tools for Fully Sharp Opinion Pooling

Consider m agents, each providing her belief as a SE about X, and a BN $\mathcal{B}$ with set of nodes $\mathbf{V}$ such that $X \in \mathbf{V}$. Straightforward application of Def. 27 would require m auxiliary nodes were added to the network. Simultaneous updating of the VEs, as by Tr. 1, would push the probability mass toward a single state. This *confirmational* effect is due to the well-known issue with posterior probability estimates in the naïve Bayes classifier [213], and might yield failure of the Consistency principle.
A least commitment approach to prevent such inconsistency would take the convex hull of all opinions from the belief profile (as done by Ex. 19); this is the *convex* PO of [238], that will be introduced in Sec. 4.2.2. In our formalism (see Ch. 3), such a CS is just the CSE $K'(X) = CH\{P_j(X) : j = 1, \dots, m\}$. Yet, given any small $\epsilon > 0$, suppose $P_1(x) = \epsilon$, $P_2(x) = 1 - \epsilon$, and $P_j(x) = p \in (\epsilon, 1 - \epsilon)$ for every $j = 3, \dots, m$, with $m \gg 3$. Despite the almost-unanimous consensus on p, taking the convex hull of the belief profile would produce a CS very close to the vacuous CS $K_0(X)$. Now, to what extent such approach should be preferred to the *confirmational* one above, say under a decision making perspective?
In this direction, a compromise solution might be offered by the following procedure:

Transformation 6. *Consider a BN over* $\mathbf{V}$ *and a collection of SEs on* $X \in \mathbf{V}$, $\mathbf{P}(X) = \{P_j(X) : j = 1, \dots, m\}$. *For each* j*, augment the BN with auxiliary binary child* $D_X^{(j)}$ *of* X*, whose CPT is such that:*

$$P(d_X^{(j)}|x) = \eta \left[\frac{P_j(x)}{P(x)} \right]^{\alpha_j}$$

with η *as in previous transformations, shared by all columns of the CPT,* $j = 1, \dots, m$.

It holds:

Proposition 4. *Consider the BN returned by Tr. 6. Then, for any target event* $\alpha \in \Sigma$*:*

$$\sum_{x \in \Omega_X} P(\alpha, x) LogOp(x) = P(\alpha | d_X^{(1)}, \dots, d_X^{(m)}) \,. \tag{4.6}$$

Our proposal, that we call *Naïve Pooling Operator* (NPO), consists in augmenting the BN as by Tr. 6, and instantiating all newly introduced auxiliary nodes to their truth values, analogously to the generalized Pearl's methods introduced in Ch. 3. By Prop. 4, NPO enhances simultaneous merging of a collection of beliefs from a pool by LogOp, and propagation of the resulting PMF through a given Bayesian Network, i.e. revision of P over Ω. Proof to the proposition is analogous to that of the first part of Prop. 2, and follows from the fact X d-separates all auxiliary nodes from the rest of the DAG.

We may now introduce the *Informed Logarithmic Pooling Operator*, extending Eq. (4.4) to the conditional framework. Such pooling process refers to context-specific beliefs such that shared awareness of the setting by the agents is incorporated by definition; see [28] for a discussion on the subject.

Definition 44 (Informed LogOp [102]). *Let m agents provide each a SE on X. Let $\Omega_{\mathbf{C}}$ be the set of all possible contexts, whose generic element is $\mathbf{c}$, such that $\mathbf{C} \subseteq \mathbf{V} \backslash \{X\}$. Suppose all agents share awareness of the current context, i.e. they all know $(\mathbf{C} = \mathbf{c})$. Also, let $\{\alpha_{\mathbf{c}} : \alpha_{\mathbf{c}} \in \mathbb{R}^m_+, \sum_{j=1}^m \alpha_{\mathbf{c},j} = 1\}$ be some collection of context-specific weights. If several conditioning events are accounted for, say $\mathbf{c}$ and $\mathbf{c}'$, for each agent j, it holds $\alpha_{\mathbf{c},j} \neq \alpha_{\mathbf{c}',j}$, $j = 1, \ldots, m$. The Informed Logarithmic Pooling Operator (iLogOp) for context $(\mathbf{C} = \mathbf{c})$ is defined as:*

$$iLogOp(P_1, \ldots, P_m; \mathbf{c})(x) = \frac{\prod_{i=1}^m P_i(x|\mathbf{c})^{\alpha_{\mathbf{c},i}}}{\sum_{x \in \Omega_X} \prod_{i=1}^m P_i(x|\mathbf{c})^{\alpha_{\mathbf{c},i}}} \tag{4.7}$$

Trivially, if $\mathbf{C} = \emptyset$, Eq. (4.7) reduces to Eq. (4.4). It may be proved iLogOp naturally enjoys all desirable properties of the latter.
Agents' weights may be regarded as bearing information on the reliability of each source [116]. A review on the approaches proposed in the literature for the choice of the weights for LogOp may be found in [74], and references therein. When available, information on each source's reliability ought to be incorporated in the pooling process. In this spirit, [203] proposed following a hierarchical approach that assumes prior distributions on the weights. Our proposal is that of deriving weights directly from a reliability measure. As simple as it may seem, we argue our approach fits well into a setup that from the application in the Appendix (Sec. 4.2.1) where beliefs serve as evidence, although probabilistic. This way, combination of beliefs tackles *credibility* outside the pool, rather than internal consensus. Note how, when context-specific beliefs are considered, reliability of an agent is expected to change across settings, unless it holds $I(X, \mathbf{C})$; in that case, $P(x|\mathbf{c}) = P(x)$ by definition, and a single α_j may be used for every agent, $j = 1, \ldots, m$. We will come back to this below.
To derive weights, we introduce the Reliability Score as measure of accuracy. It is defined as a variation of [98]'s Rank Probability Score, which in turn generalizes Brier Score [192], introduced as a reliable measure for the validation of BNs by [245].

Definition 45. *Let X be a discrete r.v. such that there exists an ordering $\succ$ over its (finite) possibility space, and let P be a PMF over Ω_X. Let $\mathbf{C} \not\ni X$ be a set of* *random variables, a collection of n observations of X and $\mathbf{C}$ is given. The Reliability Score (RS) of P is:*

$$RS_P = \sum_{\mathbf{c}\in\Omega_{\mathbf{C}}} RS_P(\mathbf{c}) = \sum_{c\in\Omega_{\mathbf{C}}} \frac{\mathbb{I}_{\mathbf{c}}}{n_{\mathbf{c}}|\Omega_X|} \sum_{l=1}^{n_{\mathbf{c}}} \sum_{j=1}^{|\Omega_X|} \left(P(x_{j,l}) - \left(\mathbb{I}_{x_{obs}=x_{j,l}} + \gamma_{x_{j,l}}\right)\right)^2 \tag{4.8}$$

with $\gamma_{x_{j,l}} = (1-\mathbb{I}_{x_{obs}=x_{j,l}})\left(1 - \frac{abs(rk(x_{obs})-rk(x_{j,l}))}{|\Omega_X|-1}\right)$. $\mathbb{I}_*$ *denotes the indicator variable, while* rk *is the rank function, mapping* Ω_X *to* $\mathbb{N}$ *such that* $rk(x_k) > rk(x_g)$ *if and only if* $x_k \succ x_g$ *in* Ω_X*.*

Context-specific reliability satisfies $0 \leq RS_P(\mathbf{c}) \leq 1$. 0 is reached if and only if all probability values are degenerate and match the observations, while 1 is reached in the extreme case of predictions being all certain over the last (first) value of the ordering while the first (last) is observed. As a remark, $\gamma_{x_{j,l}}$ equals zero whenever P is either right or completely wrong in forecasting the behavior of X, i.e. either $x_{obs} = x_{j,l}$ or $x_{obs} = x_1$ and $x_{j,l} = x_{|\Omega_X|}$, or *vice versa*. If $\gamma_{x_{j,l}}$ is set to zero for every x_j and every l, the reliability score corresponds to the generalized multivalued Brier score (MBS), that we formally define as:

$$MBS_P = \frac{1}{n|\Omega_X|} \sum_{i=1}^{n} \sum_{j=1}^{|\Omega_X|} \left(P(x_{j,i}) - \mathbb{I}_{x_{obs}=x_{j,i}}\right)^2 \tag{4.9}$$

Eq. (4.9) slightly differs from [98]'s Rank Probability score, that considers cumulative squared sums.
Let $\{RS_{P_j}(\mathbf{c}) : j = 1, \ldots, m\}$ be the reliability scores of all agents under $\mathbf{c}$, $\alpha_{\mathbf{c},j}$ may be derived as:

$$\alpha_{\mathbf{c},j} = \frac{1 - RS_j(\mathbf{c})}{\sum_{j'}(1 - RS_{j'}(\mathbf{c}))} \tag{4.10}$$

The case $RS_{P_j}(\mathbf{c}) = RS_{P_{j'}}(\mathbf{c})$ for all pairs (j, j'), yields $\alpha_{\mathbf{c},j} = 1/m$, $j' = 2, \ldots, m$.

Example 19. *Consider two agents providing their probabilistic assessment on the outcome of an incoming football match X, such that the ordered possibility space of X is $\Omega_X = \{lose, draw, win\}$. Each agent provides belief $P_j(X|\mathbf{c})$, $j = 1, 2$, within context $\mathbf{c}$:*

$$\begin{cases} P_1(win|\mathbf{c}) = 0.40 \\ P_1(draw|\mathbf{c}) = 0.35 \\ P_1(lose|\mathbf{c}) = 0.25 \end{cases}, \quad \begin{cases} P_2(win|\mathbf{c}) = 0.50 \\ P_2(draw|\mathbf{c}) = 0.30 \\ P_2(lose|\mathbf{c}) = 0.20 \end{cases}$$

By definition, $RS_{P_j}(\mathbf{c})$ discounts unreliability of wrong predictions, the closer they are to actual observations. If $(X = draw)$, $RS_{P_1}(\mathbf{c}) = 0.247$ and $RS_{P_2}(\mathbf{c}) = 0.290$. Although both predict the wrong result, Agent 1 is more balanced in forecasting

variable X's possible states, and it is more reliable than Agent 2. Note how, if $(X = win)$ *was actually observed, Agent 2 would perform best, being more accurate:* $RS_{P_1}(\mathbf{c}) = 0.272 > RS_{P_2}(\mathbf{c}) = 0.190$.
If weights were derived based on (single) match result $(X = draw)$, $\alpha_{\mathbf{c},1} = 0.5137, \alpha_{\mathbf{c},2} = 0.4863$*; if* lose *was observed* $\alpha_{\mathbf{c},1} = 0.5321, \alpha_{\mathbf{c},2} = 0.4679$.

Similarly to Tr. 6, if our model is specified as a BN over $\mathbf{V}$, with $X \in \mathbf{V}$, we may augment $\mathcal{G}$ as follows:

Transformation 7. *Consider a BN over* $\mathbf{V}$ *and a collection of CoSEs on* $X \in \mathbf{V}$, $\{P_j(X|\mathbf{c}) : j = 1, \ldots, m\}$. *For each* j, *augment the BN with auxiliary binary child* $D^{(j)}_{X|\mathbf{C}}$ *of* X *whose CPT is such that*

$$P(d_j|x, \mathbf{c}) = \eta \left(\frac{P_j(x|\mathbf{c})}{P(x|\mathbf{c})} \right)^{\alpha_{\mathbf{c},j}} \tag{4.11}$$

for every $x \in \Omega_X$ *and some fixed context* $\mathbf{c} \in \Omega_{\mathbf{C}}$. $P(d_j|x, \mathbf{c}^*) = \eta$ *whenever no opinions are expressed conditional on context* $\mathbf{c}^*$, *for all* $(x, \mathbf{c}^*) \in \Omega_X \times \{\mathbf{C} = \mathbf{c}^*\}$.

In the general case, no restrictions are posed on the relationship between X and elements of $\mathbf{C}$; for now, let them be marginally not independent, i.e. $P(x|\mathbf{c}) \neq P(x)$, $x \in \Omega_X$. We point out independence of agents' forecasts is implicitly assumed by our formalism, conditional on X and $\mathbf{C}$.

Definition 46 (Informed Pooling)**.** *Let* $\mathcal{B} = (\mathcal{G}, P)$ *be any BN, consider its augmentation as from Tr. 7. Given context* $\mathbf{C} = \mathbf{c}$, *Informed Pooling instantiates each auxiliary node* $D^{(j)}_{X|\mathbf{C}}$ *to its truth value* d_j, $j = 1, \ldots, m$, *and* $\mathbf{C}$ *to* $\mathbf{c}$. *Evidence* $\{d_1, \ldots, d_m\}$ *is propagated. (When a single context* $\mathbf{c}$ *is also observed by all agents, as a special case,* $\{d_1, \ldots, d_m, \mathbf{c}\}$ *is propagated).*

The following holds:

Proposition 5. *Informed Pooling, i.e. introduction, specification and instantiation of auxiliary nodes* $D^{(1)}_{X|\mathbf{C}}, \ldots, D^{(m)}_{X|\mathbf{C}}$ *and* $\mathbf{C}$ *as from Def. 46, reproduces simultaneous application of iLogOp and its propagation in a given BN.*

Proof. Without loss of generality, let $X = X_0$ and D_j denote $D^{(j)}_{X_0|\mathbf{C}}$; $j = 1, \ldots, m$.

Let $\alpha_{\mathbf{c}} = \{\alpha_{\mathbf{c},1}, \ldots, \alpha_{\mathbf{c},m}\}$ for any $\mathbf{c} \in \Omega_{\mathbf{C}}$. It holds:

$$\begin{aligned}
P(x_0|d_{0,1}, \ldots, d_{0,m}, \mathbf{c}) &= \frac{P(\mathbf{c})P(x_0|\mathbf{c}) \prod_{j=1}^m P(d_{0,j}|x_0, \mathbf{c})}{\sum_{x_0} P(\mathbf{c})P(x_0|\mathbf{c}) \prod_{j=1}^m P(d_{0,j}|x_0, \mathbf{c})} \\
&= \frac{P(x_0|\mathbf{c}) \prod_{j=1}^m \left(\frac{P_j(x_0|\mathbf{c})}{P(x_0|\mathbf{c})}\right)^{\alpha_{\mathbf{c},j}}}{\sum_{x_0} P(x_0|\mathbf{c}) \prod_{j=1}^m \left(\frac{P_j(x_0|\mathbf{c})}{P(x_0|\mathbf{c})}\right)^{\alpha_{c,i}}} \\
&= \frac{\prod_{j=1}^m P_j(x_0|\mathbf{c})^{\alpha_{\mathbf{c},j}}}{\sum_{x_0} \prod_{j=1}^m P_j(x_0|\mathbf{c})^{\alpha_{c,i}}} \\
&= iLogOp(P_1, \ldots, P_m; \mathbf{c})(x_0)
\end{aligned}$$

Let $(x_Q|\mathbf{C} = \mathbf{c})$ be any target event in $\Omega_{\mathbf{V}}$. Belief propagation yields:

$$\begin{aligned}
\tilde{P}(x_Q|\mathbf{c}) &= P(x_Q|\mathbf{c}, D_{0,1} = d_{0,1}, \ldots, D_{0,m} = d_{0,m}) \\
&= \sum_{x_0 \in \Omega_{X_0}} P(x_Q|x_0, \mathbf{c})P(x_0|d_{0,1}, \ldots, d_{0,m}, \mathbf{c}) \\
&= \sum_{x_0 \in \Omega_{X_0}} P(x_Q|x_0, \mathbf{c}) iLogOp(P_1, \ldots, P_m; \mathbf{c})(x_0) \\
&\neq P(x_Q|\mathbf{c})
\end{aligned}$$

unless $P_j(x_0|\mathbf{c}) = P(x_0|\mathbf{c})$, for all $j = 1, \ldots, m$.
This is true also if query events are considered not conditional on a fixed $\mathbf{c}$. Let X_0 and $\mathbf{C}$ be such that there does not exist a configuration $(x_0, \mathbf{c}) \in \Omega_{X_0} \times \Omega_{\mathbf{C}}$ such that it holds $P(x_0|\mathbf{c}) = P(x_0)$. Let the query event be x_q, and suppose also that agents provide their opinions on a collection of contexts such that there exist at least 2 elements of $\Omega_{\mathbf{C}}$ such that $P_j(x_0|\mathbf{c}) \neq P(x_0|\mathbf{c})$ for some j, $j = 1 \ldots, m$, and some $x_0 \in \Omega_{X_0}$. Let $\tilde{P}$ denote the PMF resulting from application of Informed Pooling with respect to all such contexts, it holds:

$$\tilde{P}(x_Q) = \sum_{x_0, \mathbf{c}} P(x_Q, \mathbf{c}|x_0, d_{0,1}, \ldots, d_{0,m}) iLogOp(P_1, \ldots, P_m; \mathbf{c})(x_0) \tag{4.12}$$

□

Corollary 3. *Informed Pooling is based on CoPK.*

Cor. 3 is a direct consequence of Eq. (4.12) from the proof of Prop. 5 being just Adams conditioning (cfr Ch. 2). The following results from Prop.5 as a consequence of the pooling process being reduced to an updating task on the augmented network. [12]:

[12] Analogous considerations apply to other principles satisfied by LogOp.

Corollary 4. *iLogOp is Externally Bayesian.*

Besides previous considerations on the weights, the special case of $I(X, \mathbf{C})$ requires some additional remarks. Without loss of generality, let $\mathbf{C} = \{C\}$. Introduction of m auxiliary nodes induces as many *v-structures* in $\mathcal{G}$.[13] We argue any rational agent would not account for any contexts (C), whenever the latter are irrelevant to target event (X) upstream. This way conditioning is avoided, and the procedure reduces to that outlined above by Tr. 6 and Prop. 4; that is application of standard LogOp, followed by propagation in $\mathcal{B}$.
As a final remark, note how, whenever $Pa(X) \not\supseteq \mathbf{C}$, CPT of node $D^{(j)}_{X|\mathbf{C}}$ requires a preliminary inference step to compute $P(X|\mathbf{c})$, which is not readily available unless $Pa(X) = \mathbf{C}$. Also, introduction of auxiliary nodes for Informed Pooling would make any singly connected network multiply connected and thus increase the complexity of any inferential task on the model.

Theorem 12. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be any DAG. Let $\mathcal{G}^m$ denote $\mathcal{G}$ as augmented by introduction of m auxiliary leaf nodes by Informed Pooling. Consider the following inference tasks: i) US: compute $P(x_i|\mathbf{c})$ on $\mathcal{G}$; ii) UA: compute $P(x_i|\mathbf{c})$ on $\mathcal{G}^m$; iii) UA-m: compute $P(x_i|\mathbf{c}, d_1, \ldots, d_m)$ on $\mathcal{G}^m$; for some fixed $(x_i, \mathbf{c}) \in \Omega_{X_i} \times \Omega_{\mathbf{C}}$. UA is solved by US. Also, complexity of UA-p equals that of US if $Ch(\mathbf{C}) = \{X\}$ and $Pa(\mathbf{C}) = \emptyset$ in $\mathcal{G}$.*

Proof. Consider UA. Again, let $D_{0,j} = D^{(j)}_{X_0|\mathbf{C}}$, for simplicity of notation. Nodes $\{D_{0,j} : j = 1, \ldots, m\}$ are all unobserved random variables in $\mathbf{V}^m$. Computation of $P(x_i|\mathbf{c})$ by UA is equivalent to its computation on the subgraph induced by $\mathbf{V}^m \backslash \mathbf{C} = \mathbf{V}$, that is on $\mathcal{G}$; i.e. to US. UA may be thus be solved by US in polynomial time on singly connected networks. On general networks, US may be in turn reduced to 3SAT, which is NP-hard [45].
Let $Ch(\mathbf{C}) = \{X_0, D_{0,1}, \ldots, D_{0,m}\}$ in $\mathcal{G}^m$, with $Pa(\mathbf{C}) = \cup_{X \in \mathbf{C}} Pa(X) \backslash \mathbf{C}$ and consider UA-m. It holds:

$$\begin{aligned} P(x_i|\mathbf{c}, d_{0,1}, \ldots, d_{0,m}) &= \sum_{x_0 \in \Omega_{X_0}} P(x_i, x_0|\mathbf{c}, d_{0,1}, \ldots, d_{0,m}) \\ &= \sum_{x_0 \in \Omega_{X_0}} P(x_i|x_0) P(x_0|\mathbf{c}, d_{0,1}, \ldots, d_{0,m}) \\ &= \sum_{x_0 \in \Omega_{X_0}} P(x_i|x_0) iLogOp(P_1, \ldots, P_m; \mathbf{c})(x_0) \end{aligned}$$

[13]Remember triple $< X, D, C >$ is a v-structure when it holds $I(X, C)$ and $\neg I(X, C|D)$. V-structures are graphical objects whose role is critical in BNs: instantiation of node D makes X and C dependent. See Ch. 1 for details.

The last equality holds by Prop. 5. Instantiation of $\mathbf{C}$ to $\mathbf{c}$ and of auxiliary nodes to $d_{0,j}$, $j = 1, \ldots, m$, is equivalent to straightforward elicitation of $P(X_0|\mathbf{c})$ by $iLogOp(P_1, \ldots, P_m; \mathbf{c})$. $P(X_i|\mathbf{c}, d_{0,1}, \ldots, d_{0,m})$ is thus computed by simple marginalization of X_0 on the subgraph induced by $\mathbf{V}^m \backslash \mathbf{C}$. □

As it was beyond the scope of this work, further results on the complexity of inference were not reported. Still, let us point out the following: whenever $|\mathbf{C}| = k < d = \max_{i \in \{0,\ldots,n\}} |Pa(X_i)|$ and $\neg I(X_i, C)$, augmentation of a multiply connected network does not alter its topology, provided m is reasonably smaller compared to n, as $|Pa(D_{0,j})| = k + 1 \leq d$, $j = 1, \ldots, m$. Therefore, for any $\mathcal{G}$ that is not singly connected, we do not expect introduction of auxiliary nodes to play a detrimental role on inference.

An application of Informed Pooling may be found in the following section:

An Application of Informed Pooling: SSNet and Betting Quotes

Our application is based on the BN from Fig. 4.1, called SSNet, modeling the behavior on the field of Italian football team S.S. Lazio of Rome.[14]

The model's structure was estimated based on a partially supervised search-and-score procedure. Data were collected on 45 encounters through Seasons 2016-17 and 2017-18, from the open archive of Lega Serie A Tim[15]. r.v.s were represented as either Setting, Game or Player nodes (respectively, grey, white and light blue colored in Fig. 4.1; see Table 4.1 for details. Setting nodes report information shared by all agents prior to any incoming match, while Game nodes correspond to features of the match - e.g., number of goals scored by S.S. Lazio before break, percentage of central attacking and defending actions, etc. Continuous r.v.s were discretized to avoid imposing restrictions to the graphical structure of a hybrid probabilistic graphical model, i.e. a PGM with both discrete and continuous r.v.s. The choice of quantile discretization was to some extent arbitrary, as, remarkably, the estimates we obtained were robust to other approaches. Finally, Player nodes correspond to binary r.v.s, marking presence/absence of key players on the field, and the system of play chosen by the team coach.

Informed Pooling is applied with respect to probabilistic knowledge on $X_0 = \{\text{Match}\}$, r.v. reporting the outcome of a given encounter: $\Omega_{X_0} = \{\text{Win, Draw, Lose}\}$.

[14] We do not provide details on the data examined, nor on the general learning process of SSNet, e.g. robustness of its estimates, as they were way out of the scope of this work. These may be provided to the interested reader upon request to the Author.

[15] `http://www.legaseriea.it/it/serie-a-tim/archivio`

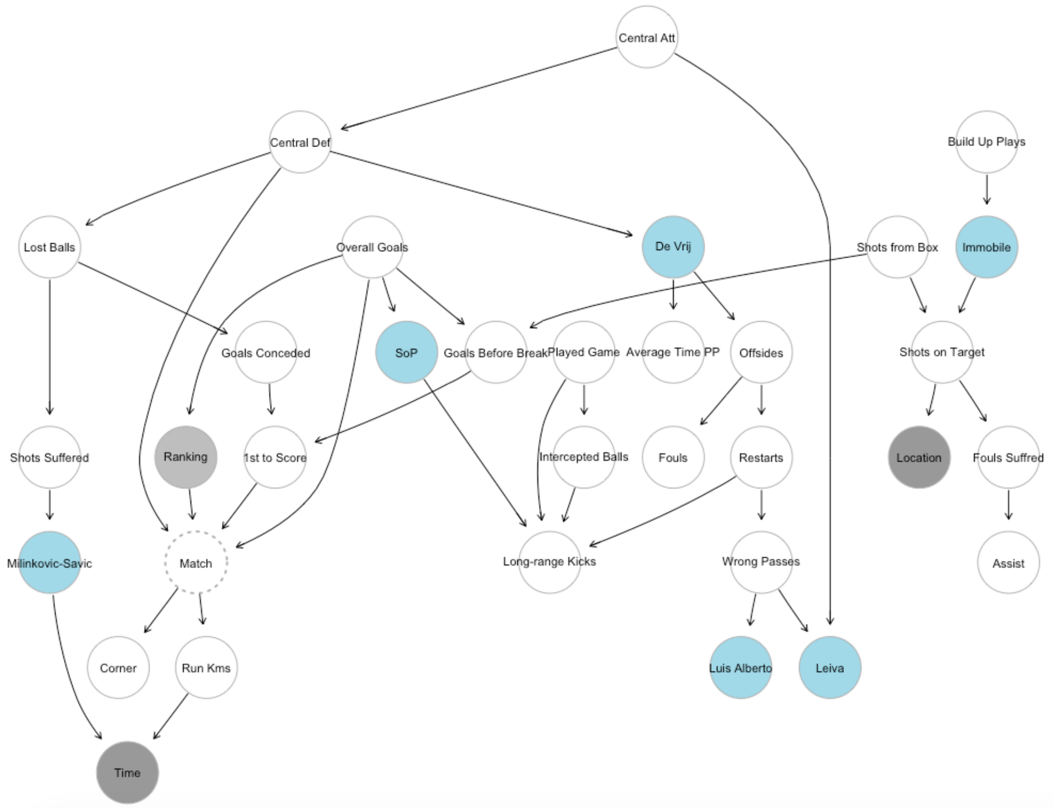


Figure 4.1: Subgraph of SS Net extended with auxiliary nodes $\text{Bet}_i, i = 1, \ldots, 4$ for informed pooling

Class	**Variables**	**Levels**
	Match (X_0)	Lose, Draw, Win
Setting	Ranking (X_1) Time (X_2) Location	Highest/High and medium/Low and lowest Regular/Prime Time/Extra Guest/Host
Game	Run Kilometers, Average Time per Player, Played Game (%), Corner Kicks, Fouls, Fouls Suffered, Restarts, Build Up Plays, Overall Shots on Target, Shots on Target From Inside the Box, Shots on Target suffered, Long-range Kicks, Wrong Passes, Assists, Offsides, Lost Balls, Central Attacking (%), Central Defending (%), Goals Conceded BB, Overall Goals, Goals Scored BB, Intercepted Balls, First to Score	Quantile discretized Yes/No
Player	De Vrij, Immobile, Leiva (Biglia) Luis Alberto (Felipe Anderson), Milinkovic-Savic System of Play (SoP)	Presence/Absence 3-5-2/4-3-3

Table 4.1: Summary of the variables used in SSNet

The intuition is the following: $m = 4$ sources for betting odds are considered, namely Bet1, Bet2, Bet3 and Bet4 [16]. From previous literature, betting companies' forecasts are recognized as most accurate, compared to any other model's [240]. We convert odds into PMFs over Ω_{X_0}, and then pool them, based on context-specific degrees of reliability of each source. Context is specified by the states of Setting nodes X_1 (Ranking) and X_2 (Time). The resulting probabilistic evidence on the upcoming match, i.e. on X_0, is used to inform SSNet about an event that is likely to happen, based on previous remarks on accuracy of forecasts.
The result of a match, given a context, is associated with some profile of the Game variables, which are in turn influenced by the asset of Player variables. Indeed, instantiation of every Player node has impact on a subset of Game variables, which in turn has an effect on X_0; we exploit such knowledge to respond to (probabilistic) evidence on X_0, and thus minimize chances of losing the game. Fig. 4.2 reports a graphical schema of the procedure outlined. We stress instantiation of Player variables ought to be considered under a sensitivity analysis approach: absence of a player is to be intended as informing the player about his average behavior on the field, that has to change in some direction (specified by SSNet), to maximize chances of winning. Note how, as a motivation to our approach, we require predictions of SSNet on X_0 to be as consistent as possible with those of betting odds: absorption of reliable knowledge is supposed to improve the performance of SSNet, not to merely replace it. This way, changes in the behavior of Game variables, related to the PMF

[16] Betting companies considered correspond to, respectively, *Better*, *bet365*, *William Hill* and *Snai*.

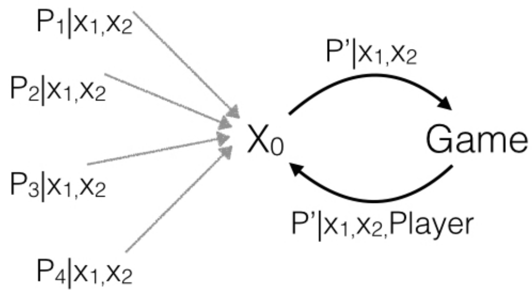


Figure 4.2: Representation of the strategic instantiation of Player nodes in SSNet. Betting companies provides PMF $P_j|x_1, x_2$, with X_1 = Ranking of the opposing team, X_2 = Time of the match. $P(\mathbf{V}|x_1, x_2)$ is revised into $P'(\mathbf{V}|x_1, x_2)$. Player nodes are instantiated to maximize $P'(X_0 = \text{Win}|x_1, x_2)$ in $\mathcal{P}_{X_0|\text{Player}}$, for each configuration of Player nodes.

over Ω_{X_0}, may be regarded as credible.

Fig. 4.3 displays the subgraph of $\mathcal{B}$ obtained by Tr. 7. Let $q_1, \ldots, q_Q$ be betting odds on, respectively, $x_1, \ldots, x_Q$[17]. Their implied probability is $p(x_q) = q_q^{-1}$, with $\sum_{q'=1}^{Q} q_{q'}^{-1} = 1 + B$, where B is known as the *bookmaker take*. We do not require $B = 0$ nor normalization of the implied probabilities since Informed Pooling, as well as iLogOp, implies normalization upstream. For a review on converting betting odds into probabilities see [240].

By Eq. (4.7), we pool forecasts, provided shared knowledge of the pool on random variables X_1 and X_2.

Each betting company provides forecasts on the result of a match, based on its accurate (yet undisclosed) system of knowledge, while pursuing its own monetary gain. We account for X_1 and X_2 to (at least partially) explain the bias of each forecast toward this latter. Briefly, we expect the ranking of the opposing team to induce a possible prejudice a betting company expresses, resulting from its own evaluation that gamblers will (not) be prone to bet against (high-) low-ranked teams. If a game takes place at regular, prime or extra time, on the other hand, possibly inflated (or deflated) odds, as well as betting flows, are expected since encounters are likely to be watched by a wider (or restricted) audience. We weight their opinions, based on our evaluation of their susceptibility to those types of effects: context-specific reliability weights of betting companies are given by Eq. (4.10).

Details on the general performance and features of SSNet, as well as those of betting companies (in terms of RS and MBS), may be found in [169]. Fig. 4.4 is reported as an example of application of Informed Pooling. It depicts the probability of the

[17] $Q = |\Omega_{X_0}| = 3$, in our application.

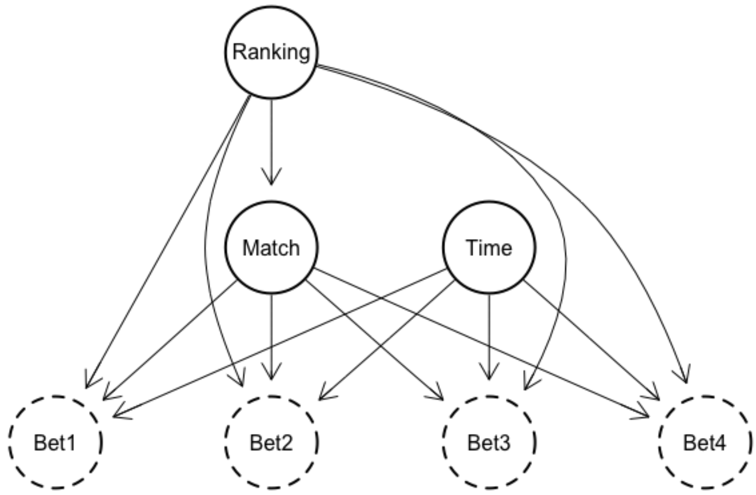


Figure 4.3: Subgraph of SS Net extended with auxiliary nodes $\mathrm{Bet}_i, i = 1, \ldots, 4$ for informed pooling

team winning the game according to the 50 best strategic choices of Player node, provided Setting nodes are fixed to i) ranking of the opposing team is *High and Medium*, ii) the match is disputed at *Extra time*, and iii) S.S. Lazio is not hosting the game.

Specialization of graphical pooling, let it be standard or Informed, to the case of pairwise mutually not independent epistemic is not straightforward. A viable approach would be that of resorting to copula graphical models [96], and take Pearson's correlation as a measure of dependence among pairs of agents, to reproduce the Copula Pooling Operator [204]. We stress Pearson's measure is tailored on linear dependence by definition, and fails in detecting asymmetric forms of dependence as well as nonlinear ones. Also, it is only appropriate in describing the strength of the dependence among a given pair when the joint opinion of the two forms an ellipse, e.g., it is normally distributed [185, 97]. Most importantly, propagation mechanics such as those proposed throughout this section do not readily apply to copula PGMs, particularly with Multinomial r.v.s. A viable approach would suggest moving from the naive Bayes classifier-like structure of Tr. 6 to a tree-augmented Bayesian network. Future work will focus on such issues.

4.2.2 Sharp-to-Imprecise Pooling

Pooling Operators of this type map a sharp belief profile to a set of PMFs in $\mathcal{P}_X$. Convexity of the set may or may not be requested, although on an epistemic stand-

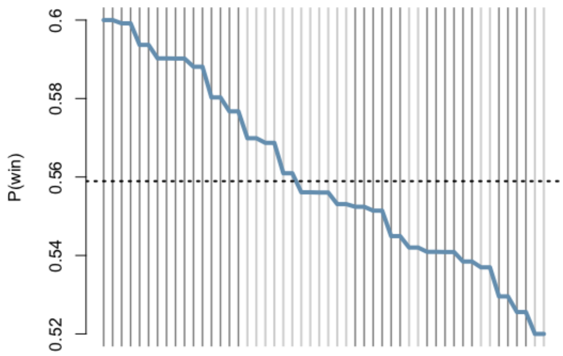


Figure 4.4: Probability of event ($X_0 =$ win), given X_1 =High or Medium, X_2=Extra time and Location=Guest, associated with the best $b = 50$ strategies, obtained by considering configurations of Player nodes. Darker vertical lines highlight Strategies with System of Play $3-5-2$. Dashed line corresponds to SSNet prior forecast.

point we argue its failure raises serious questions on the acceptability of the set itself, particularly under a Sensitivity Analysis approach. On the other hand, convex compromises are questioned within decision making setups as they are likely to imply violations of Pareto constraints on preference whenever two (or more) agents behave differently, in terms of both probability and utility [221]. A number of decision operators ignore convexity (as well as non-convexity) in the first place [257].
We introduce the convex PO, extensively discussed in [1, 238, 237]. Remarkably, in [238, 237] proofs of fulfillment of most of the principles from Sec. 4.1 did not make use of the convexity assumption, therefore enhancing the use of non-convex POs.

Definition 47 (Convex Pooling)**.** *Convex pooling maps an element from $\mathcal{P}_X^m$, to a (proper) subset in $\mathcal{P}_X$. It is generated as the convex hull induced by sharp belief profile* $\mathbf{P}$*. Formally,*

$$\Pi(X) = CH\{(P_1(x), ..., P_m(x)) : x \in \Omega_X\} .$$

Convex Pooling always satisfies Neutrality [238]. Marginalization is satisfied by any sharp-to-imprecise (not necessarily convex) PO if and only if Anonymity is also satisfied [179, Th.3.1][238, Prop.1]. Convex pooling always fulfills both [238, Prop.2]. A general characterization of the Consistency principle, that may be found in the literature as Boundedness [111] or Reasonable Range Principle [1], follows by definition from any PO taking the convex hull of a given belief profile. Additionally, External Bayesianity is satisfied by every sharp-to-imprecise PO whenever the adjustment process is based on PK or on general Imaging [237, Prop.4, Prop.7]. If the PO is that from Def. 47, convexity is also preserved. Finally, the Convex

Pooling Operatir satisfies Strong Independence Preservation by definition, whereas - unsurprisingly - Stochastic Independence Preservation is not [238].

As already mentioned, CSE (as well as CVE) may result from a collection of assessments over r.v. X, by taking their convex hull. The generalized revision rules proposed throughout Ch. 2 and 3 naturally encode Convex Pooling, and its simultaneous propagation in a BN. This finds motivation also in the fact that two equivalent sets induce the same convex hull, and thus might weaken the resistance toward the convexity requirement.

4.2.3 Imprecise-to-Sharp Pooling

Consider m imprecise beliefs on Ω_X. To avoid cumbersome notation, we use the symbol $\mathbf{K}(X)$ to denote the belief profile, although we do not make any assumptions on its elements' convexity. That is, we define each set of PMFs based on a set of (linear) constraints that define each source's belief; if convexity was assumed, we would require $K_j(X) = \{P : \underline{P}_j(x) \leq P_j(x) \leq \overline{P}_j(x), \forall x \in \Omega_X,\, j = 1, \ldots, m\}$.
Imprecise-to-sharp pooling is also referred to as *Social Inference Process* in the literature, see e.g. [264, 2]. Roughly, an Inference Process Θ searches the set of elements provided, and returns the one that maximizes a given function. As an example, the Entropy Inference Process, applied on a CS $K(X)$, identifies the PMF that maximizes Shannon's Entropy [225] as follows:

$$\Theta(K(X)) = P^{\Theta}(X) = \mathrm{argmax}_{P(X)\in K(X)} \sum_{x\in\Omega_X} P(x) \log P(x)\,.$$

When $K(X)$ is not readily available, and m imprecise beliefs are provided, a preliminary pooling process is required, whose output serves as input to Θ. The whole procedure is referred to as *Social Inference Process* [2]. This type of procedure presents some issues as there is no unique solution to the general task: let $\cap_{j=1}^{m} K_j = K^*$, either $K^* = \emptyset$ (inconsistent opinions) or $K^* \neq \emptyset$. While the first case yields no straightforward solution, the second allows every PMF in K^* to be a candidate shared agreement within the pool.
In order to properly define the general Social Inference Process sPO, let us introduce the following:

Definition 48 (Convex Bregman Divergence [29])**.** *Let f be any strictly convex function continuously differentiable, defined on a closed and convex space. $B(x, y)$ is a Convex Bregman Divergence if it is defined by the first-order Taylor approximation of $f(x)$ at y:*

$$B(x, y) = f(x) - (f(y) + < \nabla f(y), x - y >)\ .$$

By convexity of f, $B(x, y)$ is always non-negative, and zero is reached if and only if $x = y$. Let x and y be $|\Omega_X|$-dimensional vectors, it is easy to see $B(x, y) = L_2(x, y) = \|x - y\|^2$ when $f(x) = \|x\|^2$, $B(x, y) = KL(x\|y) = \sum_{i=1}^{|\Omega_X|} x_i \log \frac{x_i}{y_i}$ when $f(x) = \sum_{i=1}^{|\Omega_X|} x_i \log(x_i)$, and so forth (see [37]).

Definition 49 (sPO [264])**.** *sPO maps any set of elements of $\mathcal{P}_X^m$ into a single element of $\mathcal{P}_X$. The latter is the solution to the inference process Θ. Formally,*

$$sPO(X) = \Theta(K^B(X)) \,. \tag{4.13}$$

$K^B(X) \subseteq \mathcal{P}_X$ is defined as follows:

$$K^B(X) = \left\{ P(X) : P(X) = argmin_{P(X) \in \mathcal{P}_X} \sum_{j=1}^{m} B(P_j, P), P_j \in K_j, j = 1, \ldots, m \right\} , \tag{4.14}$$

where $B(Q, P)$ is any convex Bregman Divergence.

This class of POs fall in the class of Supra-Bayesian operators. It was proved [2] (and related works) that if the reverse KL-divergence is chosen as Bregman divergence and Θ is the Entropy Inference Process, $K^B(X)$ corresponds to the Credal LogOp (that we are going to introduce in Sec. 4.2.4), and the PMF resulting is unique. Analogously, if the KL-divergence (or the Euclidean distance) is chosen, and Θ is the Limit Centre of Mass Inference Process, defined as:

$$\Theta(K(X)) = \mathrm{argmax}_{P(X) \in K(X)} \sum_{x \in \Omega_X} \log(P(x_i)) \,,$$

(see [103] for details), $K^B(X)$ results from the Credal LinOp (Sec. 4.2.4), and the PMF resulting is unique.
sPO applies to imprecise beliefs identified by linear constraints [2, Th.1.3.1]; this is indeed the case of credal sets. Further details and results may be found in [264, 2, 103]. The general sPO[18] satisfies Anonymity, Permutation, Consistency (and Collegiality), Locality [264]. Further details may be found in the upcoming section, where the pooling task of the Social Inference Processes mentioned is considered.

4.2.4 Fully Imprecise Pooling

As discussed in Ch. 1, belief functions from Evidence theory may be regarded as CSs (remember the converse is not true in general). There, Dempster's Rule of

[18] An alternative (more general) PO was proposed in [2] as the *obdurate PO*.

Combination (DRC [224], denoted with $\oplus$) is applied to aggregate distinct (and mutually independent) beliefs on a shared domain, say Ω_X. Let $m = 2$, the pooled mass function writes:

$$(m_1 \oplus m_2)(x) = \frac{\sum_{\substack{\mathbf{x}_1,\mathbf{x}_2: \\ \mathbf{x}_1 \cap \mathbf{x}_2 = x}} m_1(\mathbf{x}_1) m_2(\mathbf{x}_2)}{1 - \sum_{\substack{\mathbf{x}_1,\mathbf{x}_2: \\ \mathbf{x}_1 \cap \mathbf{x}_2 = \emptyset}} m_1(\mathbf{x}_1) m_2(\mathbf{x}_2)}, \tag{4.15}$$

for any $x \in \Omega_X$, with $m_j(\mathbf{x}_j)$ being the mass function of source j, defined on domain $\Omega_{\mathbf{X}_j}$, $j = 1, 2$. Further combination rules were proposed for aggregating belief functions (see, e.g. [222] for a review), including the unnormalized DRC [82, 201]. See also [136, 165] for a discussion on the links between fully imprecise POs from Evidence theory and Jeffrey's rule.
An alternative approach was proposed by Capotorti *et al.* [33]. The authors focus on the coherence of an aggregated belief function, and propose merging a collection of imprecise assessments, over (possibly) overlapping domains, toward a shared consensus. This latter results as the solution to a constrained minimization problem, whose objective function is a properly defined discrepancy measure.[19] Remarkably, it satisfies the consistency principle of Unanimity Preservation, Collegiality, Equivalence and External Bayesianity. As a key feature, the proposal in [33] (and related works) is well suited to tackle inconsistencies. Also, the approach applies to generalized conditional and joint assessments. Nonetheless, it is aimed to solve a purely distance-based task, with a major focus on coherence, and no room for a discussion on conservativeness. For this reason, we would rather not stick to such proposal.
An additional approach worth mentioning is the (asymmetric) Supra-Bayesian approach, proposed by Benavoli and Antonucci [21]. Their pooling process is based on a high-level form of aggregation of belief sources, each equipped with a descriptor of state (including its reliability). Their data-fusion architecture is based on hierarchical application of the GBR (see Ch. 1), aimed at computing the posterior CS, expressed as a conditional lower prevision. The resulting probabilistic model satisfies consistency requirements when the intersection of all judgments identifies a non-empty set of PMFs. Remarkably, the reliability-based proposal in [21] was used, in combination with a properly defined supra-Bayesian (prior) belief, to deal with the well-known Zadeh's paradox [269] for DRC.

Let us go back to the Social Inference Processes, previously introduced in Sec. 4.2.3. It is easy to see that, if the first step only from the proposal in [2] is considered, the fully imprecise PO results. That is, for any Bregman divergence B, let $K^B(X)$ be

[19] Such measure was proved by the authors to be the sum of two (different) Bregman divergences, and enjoys a number of properties that makes it analogous to the latter (see [33] for an introduction) - and thus similar in spirit to the approach from [103] and related works.

the pooled imprecise opinion obtained from Eq. (4.14). It was proved $K^B(X)$ is a CS, corresponding to the set of all fixed points $\mathbf{p}^B$, obtained by applying a properly defined fully sharp PO to each configuration $(P_1, \ldots, P_m)$, for every $P_j \in K_j$, so long no vacuous judgments are provided (see [2, Ch.3] for a detailed discussion and proof of results). Their contributions provide the imprecise counterpart to Eq. (4.5):

$$K^{KL/L_2}(X) = \{\Pi(X) : \Pi(X) = LinOp_{P_1,\ldots,P_m}(X), \forall P_j \in K_j, j = 1, \ldots, m\} \ ,$$

$$K^{KL^{\leftarrow}}(X) = \{\Pi(X) : \Pi(X) = LogOp_{P_1,\ldots,P_m}(X), \forall P_j \in K_j, j = 1, \ldots, m\} \ ,$$

where $KL^{\leftarrow}$ denotes the reversed KL divergence.
As a remark, let us point out this first step of the Social Inference Process may be regarded as falling in the class of linear aggregation of beliefs - i.e. as transferable belief model -, whereas the choice of a single PMF requires a chairman, or supra-Bayesian agent, to establish an adequate criterion, e.g. maximum entropy. Furthermore, it was proved the B-based POs proposed in [2] yield the corresponding *average projection* of the pool's belief profile into its induced convex hull.
The upcoming section will introduce a graphical implementation of the revision induced by a fully imprecise (kinematical) PO on a graphical model. The operator will be referred to as Credal LogOp, although it corresponds to $K^{KL^{\leftarrow}}(X)$, which is thus motivated both under a distance-based and conservative approach.

Graphical Tools for Fully Imprecise Opinion Pooling

Now let us consider the case of m agents providing CSEs about X_n, say $\{K'_j(X_n)\}_{j=1}^m$. Let $\tilde{K}'(X_n)$ denote the CS including the output of the operator as in Eq. (4.4) for each $P'_j(X_j) \in K'_j(X_n)$ and $j = 1, \ldots, m$ [2]. Then, let us first generalize Tr. 6 as follows.

Transformation 8. *Consider a BN over* $\mathbf{X}$ *and the collection of CSEs* $\{K'_j(X_n)\}_{j=1}^m$. *For each* $j = 1, \ldots, m$, *augment the BN with a binary child* $D^{(j)}_{X_n}$ *of* X_n *whose CCPT is such that* $\underline{P}(d^{(j)}_{X_n}|x_n) \propto \left[\frac{\underline{P}'(x_n)}{P(x_n)}\right]^{\alpha_j}$ *and* $\overline{P}(d^{(j)}_{X_n}|x_n) \propto \left[\frac{\overline{P}'(x_n)}{P(x_n)}\right]^{\alpha_j}$.

This transformation returns a CN. A result analogous to Pr. 4 can be derived.

Theorem 13. *Consider the same inputs as in Tr. 8. Then:*

$$\underline{\tilde{P}}'_{X_n}(x_0) = \underline{P}(x_0|d^{(1)}_{X_n}, \ldots, d^{(m)}_{X_n}) \,, \tag{4.16}$$

where the lower probability on the left-hand side has been computed by absorption of the single CSE $\tilde{K}'(X_n)$ *and the probability on the right-hand side has been computed in the CN returned by Tr. 8. The same relation also holds for the corresponding upper probabilities.*

Proof. Consider the CN returned by Tr. 8. Pr. 4 holds for any BN consistent with the CN. Thus, the thesis just follows by taking the minimum on both sides of Eq. (4.6) with respect to the corresponding CSs. □

Corollary 5. *The simultaneous pooling and propagation of the pool's aggregated (imprecise) belief is based on CPK.*

This chapter considered pooling techniques for belief merging under generalized settings, and their implementation with BNs. Future work will extend these contributions to CNs. Next chapter will generalize the discussion to the case of uncertain instances on several variables. Such extension will require dropping the so-called *one-shot* assumption of AGM theory.

Chapter 5

Iterated Belief Revision

Throughout this thesis concepts from the symbolic theory of belief revision were deliberately confined within the probabilistic language. This way, we got around several fundamental requirements and epistemic issues that arise with general propositional logic. These were actively posed in several fields, including the philosophy of science, logic and artificial intelligence, since introduction of AGM theory in the 1980s.[1] As briefly discussed in Ch. 2, belief change theory was formulated with respect to a static setting, that we referred to as *one-shot*.
It is known by this point a doxastic agent's belief is equipped with an entrenchment ordering[2] over her beliefs, prior to revision. If such representation applies afterwards - her beliefs are equipped with an entrenchment ordering both before and after adjustment upon new information - the so-called principle of *categorical matching* is satisfied. A landmark contribution worth mentioning in this account is that of Darwiche and Pearl [65], who introduced four additional postulates for consistency of iterated belief revision.[3]
Iterated belief revision on a collection Φ of k formulae, usually accounts for an ordering, e.g. each piece of information is received by the agent in time, or it is ranked based on reliability. If this is so, previously introduced consistency requirements, i.e. retainment of new information by the agent, are now subject to the priority of each observation is equipped with. If, say, revision occurs on a time line, we are willing to retain most recent pieces of evidence, while dismissing previous revising knowledge, whenever this is contradicted by any successive observation.[4]

---

[1]The interested reader may refer, among others, to [214] for a thorough introduction to iterated belief revision (as non-monotonic reasoning).

[2]Formalized as systems of spheres, or plausibility orderings.

[3]Among others, we refer the interested reader to the work of Zaffalon and Miranda [272] on further approach to iterated belief revision with coherent lower previsions.

[4]See, e.g. [92] for a proposed systematization, and [163] for complexity results.

We argue, based on previous chapters, probabilistic iterated belief revision ought to be based on - our general account of - Iterative Probability Kinematics (Def. 51), if any ordering is known by the agent over the output of the observational process. We refer the interested reader to the axiomatic extension of PK to the iterated setting by [12], in the general framework of propositional logic. That said, our interest in this chapter revolves around adjustment of a (probabilistic) belief when pieces of information are provided on a subset of at least 2 r.v.s in $\mathbf{X}_U \subseteq \mathbf{V}$, and no ordering is available. There, revision occurs in a static setup and the only hierarchy that may be established on evidence is induced by the strength of each observation; coarsely, we expect certain findings to dominate vacuous knowledge. We shall refer to this sort of revision process as (one-step, or simultaneous) *multiple* belief change, rather than *iterated*, to avoid possible misunderstandings (and disappointment of rigorous readers!); see Fig. 5.1 for an intuition. In this framework, we account for Conservativeness postulates - as straightforward extensions of those stated for the standard case - and simultaneous full retainment of evidence. When this may not be achieved, we provide a heuristic to evaluate failure of such a generalized success postulate.
Finally, our approach to multiple belief revision specializes to the case of DAG-based PGMs. Once again, propagation of uncertain evidence is naturally intended as answering a (probabilistic) query, rather than aimed to model revision. Analogously to Ch. 2, we believe usage of kinematical approaches, as opposed to Maximum-Entropy ones, may serve as a least commitment option for such a task.

5.1 Concepts for Iterated and Multiple Belief Revision with Sharp Probabilities

Consider the case of several uncertain instances, each provided on one element from $\mathbf{X}_U$ of $\mathbf{V}$, independently of the others. Without loss of generality, we consider $\mathbf{X}_U$ partitioned into:

$$\mathbf{X}_U = \mathbf{X}_V \cup \mathbf{X}_S, \quad \mathbf{X}_V \cap \mathbf{X}_S = \emptyset\,.$$

$\mathbf{X}_V$ and $\mathbf{X}_S$ denote sets of virtual and soft instances, respectively. As hard observations are nothing but degenerate uncertain instances, we shall include them in $\mathbf{X}_V$. This choice will be motivated shortly below.
Let us begin with $\mathbf{X}_V$, and suppose $|\mathbf{X}_V| = v \geq 2$. The collection of virtual observations on $\mathbf{X}_V$ is defined as:

$$\Lambda_{\mathbf{X}_V} = \{\lambda_{X_i}, i = 1, \ldots, v\}\ .$$

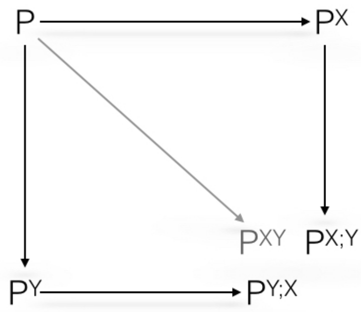


Figure 5.1: Revision schemas for $\mathbf{X}_S = \{X, Y\}$ based on orderings $\{X; Y\}$ and $\{Y; X\}$ (black), and one-step MBR (grey). Revision of PMF P yields, respectively, $P^{X;Y}, P^{Y;X}$ and P^{XY}.

By definition, multiple belief revision (MBR) on $\mathbf{X}_V$, with $|\mathbf{X}_V| = v$, leads simultaneous application of Pearl's Method:

$$P^{\circ_P}(\alpha) = \frac{\sum_{\mathbf{x}_V \in \Omega_{\mathbf{X}_V}} P(\alpha, \mathbf{x}_V) \prod_{i=1}^{v} \left(\lambda_{x_i} \mathbb{I}_{x_i \sim \mathbf{x}_V}\right)}{\sum_{\mathbf{x}_V \in \Omega_{\mathbf{X}_V}} P(\mathbf{x}_v) \prod_{i=1}^{v} \left(\lambda_{x_i} \mathbb{I}_{x_i \sim \mathbf{x}_V}\right)} .$$

In this case, it is easy to see (v-steps) iterated belief revision boils down to (one-step) MBR, since any ordering over $\mathbf{X}_V$ yields the same posterior. This is apparent if we look at the graphical representation of the method, where each D_{X_i} is separated by X_i from all other nodes in the DAG, $i = 1, \ldots, v$, and $P^{\circ_P}(\alpha) = P(\alpha | d_{X_1}, \ldots, d_{X_v})$. Things are not straightforward if belief revision is induced by two or more soft instances, i.e. whenever $|\mathbf{X}_S| = s \geq 2$. We define a collection of s soft instances as follows:

$$P'_{\mathbf{X}_S} = \left\{ P_{X_j}(X_j), j = 1, \ldots, s \right\} .$$

Iterated belief revision ought to adjust a given PMF P accordingly, provided some ordering, or *schema*, over $\mathbf{X}_S$. Let $\{X_{[1]}; \ldots; X_{[s]}\}$ be any given schema, such that r.v. $X_{[j]}$ is the j-th revising PMF, $j = 1, \ldots, s$. If SE is provided based on such an ordered sequence, e.g. based on time, revision rule $\circ$ shall be applied iteratively:

$$\left(P \circ \{P_{X_{[1]}}, \ldots, P_{X_{[s]}}\}\right)(\alpha) = \left(\left(\left(P \circ P_{X_{[1]}}\right) \circ \ldots\right) \circ P_{[s]}\right)(\alpha) ,$$

for any target event $\alpha \in \Sigma$. See Fig. 5.1 as an example, with $s = 2$, based on [255, Fig. (3.10)]. Clearly, when $P^{X;Y}, P^{Y;X}$ and P^{XY} coincide, iterated belief revision boils down to MBR, as it is for $\mathbf{X}_V$.
If SE is available on not mutually independent random variables, different PMFs shall result from different revision orderings based on some revision rule [255]. This

is due to the so-called *path-dependence* of $\circ$ considered, e.g. $\circ_J$ [133, Ch.17]. In the general case, we expect PMF $P^{\circ}(\mathbf{V})$ to retain $P'_{X_{[s]}}(X_{[s]})$ only: the last piece of information provided.
Commutativity of iterated belief revision is satisfied whenever P° is invariant to permutations in the schema. An equivalent characterization of commutativity is the following:

Definition 50 (Commutativity)**.** *Let $\mathcal{P}'_{\mathbf{X}_S}$ be the space of PMFs consistent with collection $P'_{\mathbf{X}_S}$, defined above. Commutativity is satisfied if it holds:*

$$|\cap_{X\in\mathbf{X}_S}\mathcal{P}'_X| = 1\,.$$

A weaker form of Def. 50 would require $\cap_{X\in\mathbf{X}_S}\mathcal{P}'_X \neq \emptyset$, called *cumulativity.*[5] It is easy to prove iterated updating is commutative, and so it is any iterated revision process based on virtual instances [194]: when coping with multiple virtual findings in a BN, it is sufficient to add the necessary auxiliary children, quantify their CPTs, irrespective of P, and instantiate them to their truth value. With SE, we no longer account for retainment of information as a feature of the chosen revision rule $\circ$, but as related to i) the pattern of independence described by P on its domain, and ii) the inherent nature of information.

Thorough characterization of uncertain evidence was provided in the literature. According to some authors, probabilistic assessments must be further distinguished into either fixed or not-fixed [188]. The former are produced by a source, irrespectively of its knowledge of the system[6], whereas the latter strictly depend on the source's awareness of the environment. This difference is critical to iterated belief revision, since commutativity does not apply by definition with not-fixed SE [188].
Let X and Y be any two random variables in $\mathbf{X}_S \subseteq \mathbf{V}$, such that $P'_X(X)$ and $P'_Y(X)$ are (fixed) soft assessments, respectively. We write P^X, P^Y, $P^{X;Y}$, $P^{Y;X}$ to denote the revised PMFs obtained by iterated schemas "X only","Y only","X, then Y" and "Y, then X", respectively, as in Fig. 5.1.
The iterated revision process commutes whenever X and Y are Wagner independent (WI [255]). Formally, if it holds:

$$\frac{P^X(y)}{P(y)} = \frac{P^Y(x)}{P(x)}, \quad \forall(x,y)\in\Omega_X\times\Omega_Y\,. \tag{5.1}$$

Then, $P^{X;Y}(\alpha) = P^{Y;X}(\alpha)$ for any event α. We hereby extend the equivalence result in [84] to the case of WI. This result is further generalized to the conditional (context-

[5]Cumulativity extends *simple cumulativity for probability update* [12], that trivially requires $P(\mathbf{e}|\mathbf{e}) = 1$.

[6]That is of all other elements in $\mathbf{X}_S$.

specific) setting, i.e. for CoSE, below. Beforehand, let us introduce Shogenji's Confirmational measure [228] as the ratio:

$$S(x,y) = \frac{P(x,y)}{P(x)P(y)}\,, \tag{5.2}$$

for every pair $(x,y) \in \Omega_X \times \Omega_Y$[7], with $0/0 := 0$.

Proposition 6. *Let P be any PMF on (Ω, Σ), and let X and Y be logically independent r.v.s in $\mathbf{X}_S \subseteq \mathbf{V}$. Let also $P'_X(X), P'_Y(Y)$ be any (fixed) soft findings, X and Y are WI with respect to P'_X and P'_Y if and only if it holds:*

$$\sum_{x\in\Omega_X} S(x,y)P'_X(x) = \sum_{y\in\Omega_Y} S(x,y)P'_Y(y)\,, \tag{5.3}$$

for each $x \in \Omega_X$ and $y \in \Omega_Y$.

Proof. The proof requires simple algebraic passages. Let

$$\sum_x S(x,y)P'_X(x) = \sum_y S(x,y)P'_Y(y) = k > 0\,.$$

Suppose we first revise P based on P'_X. It holds:

$$\begin{aligned} P^X(y) &= \sum_{x\in\Omega_X} P(x,y)\frac{P'_X(x)}{P(x)} \\ &= \sum_x S(x,y)P(y)P'_X(x) \\ &= kP(y) \end{aligned}$$

And thus $P^X(y)/P(y) = k$. Analogous reasoning yields $P^Y(x) = kP(x)$. □

By Prop. 6, it holds:

$$\left(P \circ_J P_X\right)(y) = \sum_{x\in\Omega_X} P(x,y)\frac{P'_X(x)}{P(x)} = kP(y), \quad \forall y \in \Omega_Y\,,$$

and $\left(P \circ_J P_Y\right)(x) = kP(x)$, for each $x \in \Omega_X$.
WI extends stochastic and Jeffrey independence [84]. It also implies convergence in one step of any iterated approach to belief revision.[8]
A least commitment kinematical approach to iterated belief revision would require the following to hold:

[7] Remember from Ch. 1 a pair is defined on the Cartesian product space of the associated possibility spaces if and only if P the two are logically independent.

[8] See e.g. [197, 244].

Definition 51 (Iterative PK). *Let P be a PMF over (Ω, Σ), and $\mathbf{X}_S$ be an ordered collection of s random variables, such that SE is provided on every $X \in \mathbf{X}_S \subseteq \mathbf{V}$. Any iterated revision process that produces PMF P° is based on Iterated PK (IPK) if and only if, for any $\alpha \in \Sigma$, it holds:*

IPK1 $P^\circ(\alpha|\mathbf{x}_S) = P(\alpha|\mathbf{x}_S)$, *for each* $\mathbf{x}_S \in \Omega_{\mathbf{X}_S}$ *, (Iterated Conservativeness)*

IPK2 $P^\circ(X_{[s]}) = P_{X_{[s]}}(X_{[s]})$*. (Prioritized Responsiveness)*

The Success postulate is concerned with the last piece of information, with no requirements on retainment of all evidence of $P_{\mathbf{X}_S}$. As already discussed, commutativity (and thus full success of revision) comes as a feature of P on $\mathbf{X}_S$ and of SE, rather than $\circ$'s.
It is easy to prove iterated belief revision by a kinematical rule based on WI r.v.s is always based on IPK, while satisfying the stronger postulate:

IPK2' $P^\circ(X) = P'_X(X)$, *for each* $X \in \mathbf{X}_S$*. (Iterated Responsiveness)*

In the general case, we might be interested in measuring the extent to which postulate IPK2' is failed by an IPK-based rule, by computing (for any distance measure d):

$$d(P^\circ(\mathbf{X}_S \backslash \{X_{[s]}\}), P^{\mathbf{X}_S \backslash \{X_{[s]}\}}) \,, \tag{5.4}$$

where $P^{\mathbf{X}_S \backslash \{X_{[s]}\}}(\mathbf{x}_S^{-[s]}) = \prod_{X \in \mathbf{X}_S \backslash \{X_{[s]}\}} P'_X(x) \mathbb{I}_{x \sim \mathbf{x}_S^{-[s]}}$, for each $\mathbf{x}_S^{-[s]} \in \times_{X \in \mathbf{X}_S \backslash \{X_{[s]}\}} \Omega_X$. We shall return to Eq. (5.4) below.

5.1.1 The Context-Specific Case

Let $\mathbf{C}_{\mathbf{X}_S}$ be a given union set of *context variables* for $\mathbf{X}_S$, such that CoSE is provided over $\Omega_X \times \{\mathbf{C}_X = \mathbf{c}^*_X\}$, for some $\mathbf{c}^*_X \in \Omega_{\mathbf{C}_X}$, for each $X \in \mathbf{X}_S$. Analogously, $\mathbf{c}^*_{\mathbf{X}_S} = \{\mathbf{c}^*_X : X \in \mathbf{X}_S\}$ is the collection of all *relevant contexts*; as a remark, $\mathbf{c}^*_{\mathbf{X}_S}$ is not necessarily a (consistent) configuration in $\Omega_{\mathbf{C}_{\mathbf{X}_S}}$.[9]
Let us denote the collection of all conditional instances with $P_{\mathbf{X}_S|\mathbf{c}^*_{\mathbf{X}_S}}$:

$$P_{\mathbf{X}_S|\mathbf{c}^*_{\mathbf{X}_S}} = \left\{ P_{X|\mathbf{c}^*_X}(X|\mathbf{c}^*_X), X \in \mathbf{X}_S, \mathbf{c}^*_X \in \mathbf{c}^*_{\mathbf{X}_S} \right\} \,.$$

The example below shows full retainment of all pieces of information is verified, by a CoPK-based $\circ_A$, if all elements of $\mathbf{X}_S$ are mutually CSI (cfr Def. 2, Ch. 1).

[9] e.g. $\mathbf{C}_X = \mathbf{C}_Y$, yet $\mathbf{c}^*_X \neq \mathbf{c}^*_Y$, for some $X, Y \in \mathbf{X}_S$.

Example 20. *Let $X, Y, Z \in \mathbf{V}$, and $\mathbf{X}_S = \{X, Y\}$, such that $\{\mathbf{c}^*_X = \mathbf{c}^*_Y\} = \{z^*\}$. If X and Y are CSI conditional on z^*, it holds $P(x, y|z^*) = P(x|z^*)P(y|z^*)$, for each $(x, y) \in \Omega_X \times \Omega_Y \times \{Z = z^*\}$, whereas $P(x, y|z) \neq P(x|z)P(y|z)$ in general, i.e. for each $(x, y, z) \in \Omega_X \times \Omega_Y \times \Omega_Z$.*[10]
Let $P'_X(X|z^*), P'_Y(Y|z^*)$ be CoSE on X and Y, respectively. $P(y)$ is revised by Adams conditioning as follows:

$$\begin{aligned}
P^X(y) &= P(y, \neg z^*) + \sum_x P(y, x, z^*) \frac{P'_X(x|z^*)}{P(x|z^*)} \\
&= P(y, \neg z^*) + P(z^*)P(y|z^*) \sum_x P'_X(x|z^*) \\
&= \sum_z P(y, z) \\
&= P(y)
\end{aligned}$$

whereas

$$P^X(y) = P(y, \neg z) + \sum_x P(y, x, z) \frac{P'_X(x|z)}{P(x|z)} \neq P(y)$$

in the general case. It is easy to check:

$$\begin{aligned}
\left(P \circ_A \{P'_X, P'_Y\}\right)(\alpha) &= \left(P \circ_A \{P'_Y, P'_X\}\right)(\alpha) \\
&= P(\alpha, \neg z^*) + \sum_{x,y} P(\alpha, x, y, z^*) \frac{P'_X(x|z^*)}{P(x|z^*)} \frac{P'_Y(y|z^*)}{P(y|z^*)}.
\end{aligned}$$

We shall now define partial contexts as $\mathbf{c}^*_I = \{\mathbf{c}^*_i \sim \mathbf{c}^*_{\mathbf{X}_S} : i \in I\}$ and, conversely, $\mathbf{c}^*_{\neg I} = \{\mathbf{c}^*_i \sim \mathbf{c}^*_{\mathbf{X}_S} : i \notin I\}$, for some $I \subset \{\emptyset, 1, ..., s'\}$ indexing the elements of $\mathbf{X}_S$, provided $|\mathbf{X}_S| = s' \geq 2$. Every configuration in

$$\mathbf{c}^*_{\neg \mathbf{X}_S} = \Omega_{\mathbf{X}_S} \setminus \left[\cup_{I \subseteq \{\emptyset, 1, ..., |\mathbf{X}_S|\}} \overline{\mathbf{c}}^*_I\right]$$

is *neutral* with respect to $P_{\mathbf{X}_S|\mathbf{c}^*_{\mathbf{X}_S}}$ as a whole.
For a fixed I, $\mathbf{c}_{I,\neg I} = (\mathbf{c}^*_I, \neg \mathbf{c}^*_{\neg I})$ is the set of configurations in $\Omega_{\mathbf{C}_S}$ such that a) the subset of elements indexed by I is relevant with respect to the CoSE on the variables in $\mathbf{X}_I \subseteq \mathbf{X}_S$, while b) the remaining are neutral to $\mathbf{X}_I$. It follows $\mathbf{c}^*_{\mathbf{X}_S} = (\mathbf{c}^*_{\{1, ..., |\Omega_{\mathbf{X}_S}|\}}, \emptyset)$, and *vice versa* for the negated element, if they exist. Finally, let $\mathbf{c}^\alpha = \{\mathbf{c}_{I,\neg I} \mathbb{I}_{\mathbf{c}_{I,\neg I} \sim \alpha}, I \subseteq \{\emptyset, 1, \ldots, |\Omega_{\mathbf{X}_S}\}\}$, for any target event $\alpha \in \Sigma$.
If CoSE is provided on mutually CSI random variables, generalization of CoPK to the iterated setting is straightforward (as shown by Ex. 21). Conditioning must be

[10] We are always assuming $P(z) > 0$. Again, this is implied by the logical independence of the triple (X, Y, Z).

considered on all relevant and neutral configurations (respectively, for CoPK1 and CoPK2 of Def. 17) and on the union of all (partially) relevant configurations and corresponding elements of $\mathbf{X}_S$ (CoPK4).
We generalize the previous result on Shogenji's measure to the current setting. Our results apply without loss of generality to the full conditional setting, where all contexts are relevant.[11]

Definition 52 (Context-Specific Shogenji's Measure)**.** *Let $\mathbf{c}^*_X, \mathbf{c}^*_Y$ be relevant contexts for variables X and Y, respectively. Context-specific Shogenji's measure extends Eq.* (5.2) *as:*

$$S(x, y; \mathbf{c}^*_X, \mathbf{c}^*_Y) = \frac{P(x, y, \mathbf{c}^*_X, \mathbf{c}^*_Y)}{P(x, \mathbf{c}^*_X)P(y, \mathbf{c}^*_Y)}\,, \tag{5.5}$$

for all pairs $(x, y) \in \Omega_X \times \Omega_Y \times \{\mathbf{C}^*_X = \mathbf{c}^*_X\} \times \{\mathbf{C}^*_Y = \mathbf{c}^*_Y\}$.

If X and Y are CSI with respect to context $(\mathbf{c}^*_X, \mathbf{c}^*_Y)$, then $S(x, y; \mathbf{c}^*_X, \mathbf{c}^*_Y) = 1$, for all $(x, y) \in \Omega_X \times \Omega_Y \times \{\mathbf{C}^*_X = \mathbf{c}^*_X\} \times \{\mathbf{C}^*_Y = \mathbf{c}^*_Y\}$.

Proposition 7. *Let X and Y be two random variables such that CoSE $P'_X(X|\mathbf{c}^*_X)$ and $P'_Y(Y|\mathbf{c}^*_Y)$ is provided. X and Y are context-specific WI (CS-WI) conditional on relevant contexts $\mathbf{c}^*_X, \mathbf{c}^*_Y$, and with respect to the provided CoSE, if and only if it holds:*

$$\sum_{x,\mathbf{c}_X} S(x, y; \mathbf{c}^*_X, \mathbf{c}^*_Y)P^X(x|\mathbf{c}^*_X)P(\mathbf{c}^*_X) = \sum_{y,\mathbf{c}_Y} S(x, y; \mathbf{c}^*_X, \mathbf{c}^*_Y)P^Y(y|\mathbf{c}^*_Y)P(\mathbf{c}^*_Y) \tag{5.6}$$

for all $(x, y) \in \Omega_X \times \Omega_Y \times \{\mathbf{C}_X = \mathbf{c}^*_X\} \times \{\mathbf{C}_Y = \mathbf{c}^*_Y\}$, with $P(\mathbf{c}^*_X, \mathbf{c}^*_Y) > 0$.

Proof. Consider revision of the PMF by $P^X(X|\mathbf{c}^*_X)$ and let Eq. (5.6) equal $k > 0$. It holds:

$$\begin{aligned}
P^*_X(y, \mathbf{c}^*_Y) &= P(y, \mathbf{c}^*_Y, \neg\mathbf{c}^*_X) + \sum_x P(x, y, \mathbf{c}^*_X, \mathbf{c}^*_Y)\frac{P^X(x|\mathbf{c}^*_X)}{P(x|\mathbf{c}^*_X)} \\
&= \sum_x \left[P(y, x, \mathbf{c}^*_Y, \neg\mathbf{c}^*_X) + S(x, y; \mathbf{c}^*_X, \mathbf{c}^*_Y)P(y, \mathbf{c}^*_Y)P(\mathbf{c}^*_X)P^X(x|\mathbf{c}^*_X)\right] \\
&= P(y, \mathbf{c}^*_Y)\sum_{x,\mathbf{c}_X} S(x, y; \mathbf{c}_X, \mathbf{c}^*_Y)P^X(x|\mathbf{c}_X)P(\mathbf{c}_X) \\
&= P(y, \mathbf{c}^*_Y)k\,,
\end{aligned}$$

provided $P^X(x|\mathbf{c}_X) = P(x|\mathbf{c}_X)$ when $\mathbf{c}_X \neq \mathbf{c}_X$.
Conversely,

$$\sum_{x,\mathbf{c}_X} \frac{P(x, y, \mathbf{c}_X, \mathbf{c}^*_Y)}{P(x|\mathbf{c}_X)}P^X(x|\mathbf{c}_X)P(\mathbf{c}^*_X) = kP(y, \mathbf{c}^*_Y)$$

[11] I.e. $\mathbf{c}^*_{\mathbf{X}_S} = \Omega_{\mathbf{X}_S}$.

if and only if it holds

$$\sum_{x,\mathbf{c}_X} \frac{P(x,y,\mathbf{c}_X,\mathbf{c}_Y^*)}{P(x,\mathbf{c}_X)P(y,\mathbf{c}_Y^*)} P^X(x|\mathbf{c}_X) = \sum_{x,\mathbf{c}_X} S(x,y;\mathbf{c}_X,\mathbf{c}_Y^*)P^X(x|\mathbf{c}_X)P(\mathbf{c}_X) = k$$

Results for $(X,\mathbf{c}_X)$ may be derived by analogous reasoning. □

We shall formally define Adams conditioning's extension to simultaneous MBR:

Definition 53 (Multiple Adams Conditioning)**.** *Let $\mathbf{X}_S$ be a collection of mutually CS-WI random variables on which CoSE is provided and let α be any target event. $P^*(\alpha)$ is obtained by Multiple Adams Conditioning (MA) if it holds:*

$$P^*(\alpha) = P(\alpha,\neg\mathbf{c}^{\alpha}_{\mathbf{X}_S}) + \sum_{I\subseteq\{1,\ldots,k\}} \sum_{\mathbf{x}_I\in\Omega_{\mathbf{X}_I}} P(\alpha,\mathbf{x}_I,\mathbf{c}^{\alpha}_{I,\neg I}) \prod_{X\in\mathbf{X}_I} \frac{P'(x|\mathbf{c}_X^*)}{P(x|\mathbf{c}_X^*)}, \qquad (5.7)$$

where $\neg\mathbf{c}^{\alpha}_{\mathbf{X}_S}$ is the set of all contexts neutral to $P_{\mathbf{X}_S|\mathbf{c}^*_{\mathbf{X}_S}}$ consistent with α. We write $P^*(\alpha) = \left(P \circ_{MA} P_{\mathbf{X}_S|\mathbf{c}^{\alpha}}\right)(\alpha)$, and denote the MA revision of P with $P^{\circ_{MA}}$.

Let $\mathbf{C}_X, \mathbf{C}_Y$ be two sets of context r.v.s for $X, Y \in \mathbf{X}_S$, respectively. Suppose $\mathbf{C}_X \cap \mathbf{C}_Y = \mathbf{A} \neq \emptyset$ (as in Ex. 21). Without loss of generality, suppose $\mathbf{c}_X^* = \mathbf{c}_Y^* = \mathbf{a}^*$. When $k = 2$, Eq. (5.7) reduces to

$$P'(\alpha) = P(\alpha,\neg\mathbf{a}^*) + P_J(\alpha,\mathbf{a}^*) \qquad (5.8)$$

The second term of the right-hand side of Eq. (5.8) corresponds to Jeffrey's rule applied on the $\mathbf{a}^*$-slice of $\Omega_X \times \Omega_Y \times \Omega_{\mathbf{A}}$ [12].

A routine for the detection of CSI relationships in a partially oriented DAG is presented in the Appendix (Sec. 2). Results from some applications are also reported. The procedure presented shall be further applied to either i) deal with faithfulness violations (or quasi-violations), or ii) reduce parametric dimensionality (we provide results on benchmark networks from the literature).

Next section extends results from Ch. 3 to the iterated and multiple setups. Credal instances are also considered therein.

[12] That is

$$\begin{aligned}
P_J(\alpha,\mathbf{a}^*) &= \sum_{y\in\Omega_Y}\sum_{x\in\Omega_X} P(\alpha,x,y,\mathbf{a}^*)\frac{P'(x|\mathbf{a}^*)}{P(x|\mathbf{a}^*)}\frac{P'(y|\mathbf{a}^*)}{P(y|\mathbf{a}^*)}\\
&= \sum_{x,y} P(\alpha,x,y,\mathbf{a}')\frac{P'(x,y|\mathbf{a}^*)}{P(x,y|\mathbf{a}^*)}\\
&= \sum_{x,y} P(\alpha|x,y,\mathbf{a}^*)P'(x,y,\mathbf{a}^*)\\
&= \left((P \circ_J P_X) \circ_J P_Y\right)(\alpha,\mathbf{a}^*)
\end{aligned}$$

The second equality follows from CSI of X and Y with respect to $\mathbf{a}^*$. Also, by CSI it holds $\left((P \circ_J P_X) \circ_J P_Y\right)(\alpha,\mathbf{a}^*) = \left((P \circ_J P_Y) \circ_J P_X\right)(\alpha,\mathbf{a}^*)$.

5.2 Related Work on Iterated Belief Revision

Previous works on iterated belief revision in the general setting of jointly inconsistent soft instances on $\mathbf{X}_S$ stem from the properties of standard constrained IPFP. As already discussed, IPFP for belief revision leads (in the limit) the unique KL-projection [13] of the prior PMF on the space of probability distributions that are consistent with SE on $\mathbf{X}_S$, while minimizing the KL divergence (and total variation) [197, Sec.2]. We briefly outline the procedure with Jeffrey's rule. Let $|\mathbf{X}_S| = s$, $s \geq 1$, IPFP iteratively applies the following, until convergence is reached:

$$\begin{cases} P_{(0)}(\mathbf{v}) = P(\mathbf{v}) \\ P_{(i)}(\mathbf{v}) = \left(P_{(i-1)} \circ_J P_{X_j}\right)(\mathbf{v}) \quad j = ((i-1)\mathrm{mod}(s)) + 1 \end{cases} \tag{5.9}$$

with $P_{(i)}(\mathbf{v}) = 0$ by definition, if $P_{(i-1)}(x_j) = 0$ at some point.
When $s = 1$, Eq. (5.9) trivially reduces to Jeffrey's rule, and IPFP halts after the first step, with $(j = 1, i = 1)$. Analogously, if all variables are mutually WI, the procedure converges at $(j = s, i = 1)$. This is equivalent to application of Jeffrey's rule on joint r.v. $\mathbf{X}_S$ in a single step; i.e. conditioning on joint event $(\mathbf{X}_S = \mathbf{x}_S)$, with $P(\mathbf{x}_S) = \prod_{j=1}^{k} P(x_j)$, for each $\mathbf{x}_S \in \Omega_{\mathbf{X}_S}$.[14]
When cumulativity does not hold, a naïve approach to iterated belief revision would suggest to separately revise P by P_X, for each $X \in \mathbf{X}_S$. For a fixed α, s Jeffrey revisions $P^{\circ_J,1}(\alpha), \ldots, P^{\circ_J,s}(\alpha)$ result. Then, take $s^{-1} \sum_{j=1}^{s} P^{\circ_J,j}(\alpha)$ as the solution. Vomlel proposed in [250] an EM-based routine for (approximate) iterated belief revision, called GEMA (Generalized EM Algorithm), that is based on the naïve intuition above. GEMA imposes a priority ordering upon the elements of $\mathbf{X}_S$, via the convex combination[15] of weights $w_j, j = 1, \ldots, s$. At each iteration, the naïve method is applied, and the weighted mean is accounted for as the solution, while checking on convergence, i.e. minimization of:

$$\sum_{j=1}^{s} w_j KL(P'_{X_j} \| P_{(i)}^{\downarrow X_j}) = \sum_{j=1}^{s} w_j \sum_{x_j \in \Omega_{X_j}} P'_{X_j}(x_j) \log \frac{P'_{X_j}(x_j)}{P_{(i)}(x_j)} .$$

Among others, Peng [197] reported complexity of GEMA strongly depends on the inherent nature of P and $\mathbf{V}$.
The same author of GEMA had previously proposed, in [140], approximate serial relaxation of IPFP, where the second line of Eq. (5.9) was replaced by:

$$P_{(i)}(\mathbf{v}) = (1 - \rho_i) P_{(i-1)}(\mathbf{v}) + \rho_i \left(P_{(i-1)} \circ_J P_{X_j}\right)(\mathbf{v}) ,$$

[13] See [57], and Sec. 4.2.4.
[14] See [44] for a complete introduction to optimization when jointly inconsistent constraints are given.
[15] $w_j > 0, \sum_{j=1}^{s} w_j = 1$.

with $\rho_i \in (0,1)$. When cumulativity does not hold (nor commutativity), Vomlel's proposal converges as $\rho_1 \approx 1$ decreases to 0 with iterations - as i grows. The decreasing rate serves as tradeoff between computational feasibility and responsiveness of the (approximate) revision process.[16]

An alternative application was proposed in [197] at the constraints level, rather then on the joint PMF. We hereby sketch the procedure: at first, standard IPFP is run and either i) some revision is obtained as output, or ii) the process gets stuck looping. In the first case, cumulativity holds, and the optimal solution results from convergence of IPFP. If ii) is the case, $P^{(0)}_{X_j}(X_j)$ is initialized as $P'_{X_j}(X_j)$, $j = 1, \ldots, s$, and the following applies at each iteration k, until convergence:

$$\begin{cases} P^{(l)}_{X_j}(x_j) = \tau P^{(l-1)}_{X_j}(x_j) + (1-\tau) P_{(k-1)}(x_j), \quad \forall x_j \in \Omega_{X_j} \\ P_{(k)}(\mathbf{v}) = \left(P_{(k-1)} \circ_J P^{(l)}_{X_j} \right)(\mathbf{v}) \end{cases}$$

with $j = (1 + (k-1) \mod (s))$ and $l = (1 + \lfloor (k-1)/s \rfloor)$. In words, at $l = 1$, a mixture of the original constraint and of the evaluation of X_j by the current PMF replaces the former, and *smoothed* SE is used for revision. As the procedure iterates, such mixture-constraints move toward a feasible region for convergence. If $s = 2$, this is always reached [197, Th.7].[17]

Simply put, when both responsiveness and conservativeness can not be satisfied, distance based approaches, such as Maximum-Entropy IPFP, favor the first as opposed to kinematical methods, whose focus is on the second principle. We argue PK (and related extensions) represent a safe approach to belief adjustment, whose importance is apparent when revision is applied to probabilistic graphical models. Iterated propagation of probabilistic beliefs with DAG-based models was considered in [244, 197] and related works. Valtorta *et al.* [244] proposed application of standard IPFP among r.v.s in the root clique of a properly built junction tree. Their procedure, called *Big-clique* algorithm, accounts for the constrained construction of such tree, whose treewidth is expected to be at least s, since all r.v.s in $\mathbf{X}_S$ are in the root.[18] Efficiency of the Big-Clique algorithm may be easily improved, based on [141, 142].

Based on [35], Peng proposed replacement of SEs by VEs, as we did in Ch. 2. Remarkably, his proposals for (approximate) belief propagation in BNs all tried to patch up inconsistencies produced by failure of conservativeness in the model, via a

[16]When cumulativity does hold, the proposed serial relaxation of IPFP converges if $\rho_i = \rho$, for each i.

[17]See [197] for further details, e.g. on the optimal choice for smoothing constant $\tau \in (0,1)$.

[18]Remember from Ch. 1 the treewidth of a BN may be equivalently defined as the minimal size of the largest clique in any triangulation, and that complexity of inference in a junction tree is exponential in its width. See [197] for experimental evaluations and comparisons on complexity.

further step, invoked to formally preserve the structure of $\mathcal{G}$ (at the parents' level). We argue this does not prevent failure of conservativeness, and thus exposes the model to faithfulness violations.

5.3 Graphical Tools for Iterated and Multiple Belief Revision

We call *uncertain credal updating* (UCU) of a BN the task of computing updated/revised beliefs in a BN with an arbitrary number of CSEs, CVEs, and hard evidences as well.
UCU of a BN when VE (or conditional VE) instances only are provided is straightforward, and invariant to any revision schema.
Let us first consider the case of sharp probabilistic instances, that is SE and CoSE over $k < (n+1)$ r.v.s: $\mathbf{X}_U = \mathbf{X}_S$. If an ordering is given over the k r.v.s, iterated belief revision might be applied as a straightforward extension of our proposals from Ch. 3, based on iterative PK (Def. 51). Clearly, we expect the process to satisfy IPK2' - that is, to fully retain all instances - if and only if all pairs in $\mathbf{X}_S$ are mutually WI (or CS-WI).
Based on results of Sec. 5.1, we outline a simple proposal for kinematical MBR based on marginal and conditional soft evidence on k r.v.s in a BN, when no ordering is available for revision. This setup is generalized by the one where further virtual instances are available: these latter are internally invariant to any revision schema, and shall be treated as a single (additional) element in the iterated UCU task.

1. Augment $\mathcal{G}$ with k auxiliary nodes, as from Ch. 3. This step requires computation of Denote the resulting DAG with $\mathcal{G}^*$;

2. Build a junction tree from $\mathcal{G}^*$;

3. Instantiate all auxiliary nodes to their truth value and propagate.

Some remarks are due. As a first, if all the k r.v.s are Wagner (context-specific) independent, we expect MBR and iterated technique to yield the same revision of P. Also, without loss of generality let $\alpha = (X_q = x_q)$; if the set of all uncertain instances contains $k' \leq k$ elements that are m-irrelevant[19] to X_q (or, possibly, to x_q), MBR outlined above shall be restricted to propagation of knowledge of the remaining $k - k'$ r.v.s. This is motivated by the fact that propagation of uncertain beliefs

[19]See Def. 10.

requires marginalization of each X_i on which SE (or CoSE) is provided. If X_i is m-irrelevant to X_q, changes in its marginal (or conditional) PMF do not influence X_q's behavior, by definition. Finally, a sensitivity analysis approach, when no ordering is available, would suggest computing all revisions of $P(x_q)$, and taking its mean values and range of variation. Such sensitivity analysis approach may be applied to the case of J subsets of mutually WI (or CS-WI) r.v.s, yielding a partition of $\mathbf{X}_S$. In this case, iterated belief revision may be applied with respect to each subset $\mathbf{X}_{S,j}$, while MBR is performed at each step j, based on its k_j members, $j = 1, \ldots, J$; with $\sum_{j=1}^{J} k_j = k$. Such sensitivity analysis approach would be based on a smaller number of revision schemas.

As for credal instances and BNs, when $v \geq 2$ CVEs only are available - i.e. $\mathbf{X}_U = \mathbf{X}_V$ - we augment the DAG with v auxiliary children as in Ch. 3 and apply, e.g. ApproxLP.
The procedure becomes less straightforward when coping with multiple soft observations. As previously discussed, specification of each auxiliary child's CPT requires a preliminary inference step, and, as a consequence, iterated updating of multiple SEs might be not invariant with respect to the revision process scheduling [254], unless WI holds. Additionally, with CSEs, absorption of the first CSE transforms the BN into a CN, and the absorption of other CSEs requires a further extension of the procedure, that we consider in details below. By MBR, multiple CSEs shall be converted in CVEs and the inferences required for the quantification of the auxiliary children performed in the original BN. This trivially corresponds to a single iteration of Peng's proposal in [197], extended to the credal framework.

As for complexity of inference, with CNs, binary polytrees may be updated efficiently, while updating ternary polytrees is already NP-hard.[20] An important question is therefore whether or not a similar situation holds for UCU in BNs. The (positive) answer is provided by the two following results.

Proposition 8. *When a single query variable is considered, UCU of polytree-shaped binary BNs can be solved in polynomial time.*

Proof. The proof of this proposition is trivial and simply follows from the fact that the auxiliary nodes required to model CVE and/or CSE are binary (remember that CSs over binary variables are always shady). The CN solving the UCU is therefore a binary polytree which can be updated by the exact algorithm proposed in [100]. □

Theorem 14. *UCU of non-binary polytree-shaped BNs is NP-hard.*

[20]This is known from Ch. 1.

Proof. The proof of this theorem is based on a reduction to the analogous result for CNs [175]. Notably, this already concerns models whose variables have no more than three states and treewidth equal to two.

To prove the theorem we show that the non-binary polytree-shaped CN used by [175, Th. 1] to prove the NP-hardness of non-binary credal polytrees can be used to model UCU in a non-binary polytree-shaped BN. To do that for an arbitrary k, consider the BN over $\mathbf{X} := (X_0, X_1, \ldots, X_{2k})$ with the topology in Fig. 5.2, Nodes $(X_0, \ldots, X_{k=1})$ are associated to binary variables, the others to ternary variables. A uniform marginal PMF is specified for X_k, while the CPTs for the other ternary variables are as indicated in Table 2 of the proof we refer to (the numerical values being irrelevant for the present proof). For the binary variables we also specify a uniform prior.

We specify indeed a *vacuous* CSE for each binary variable. These CSEs can be asborbed by replacing the uniform PMFs with vacuous CSs. The resulting model is exactly the CN used to reduce CN updating to the *PARTITION* problem [110] and hence proves the theorem.

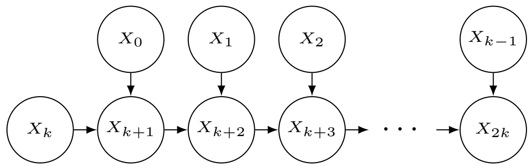


Figure 5.2: A polytree-shaped directed acyclic graph.

□

We provided some theoretical results on the independence conditions that allow for commutativity of the iterated belief change process, with sharp probabilities. Also, we extended previous contributions, on the graphical implementation of our proposed operators for kinematical belief propagation under uncertainty and imprecisions, to the case of several instances. Future developments will be focused on the fully general case, of credal instances to be propagated in CNs.

Conclusions

The present work stands on the border between information theory, artificial intelligence, and probabilistic and statistical theory. We aimed to introduce a unified treatment of uncertain and unreliable evidence in complex probabilistic systems of knowledge, laying the foundations of a sound approach to radically different literatures, each equipped with its own language and properties. These are:

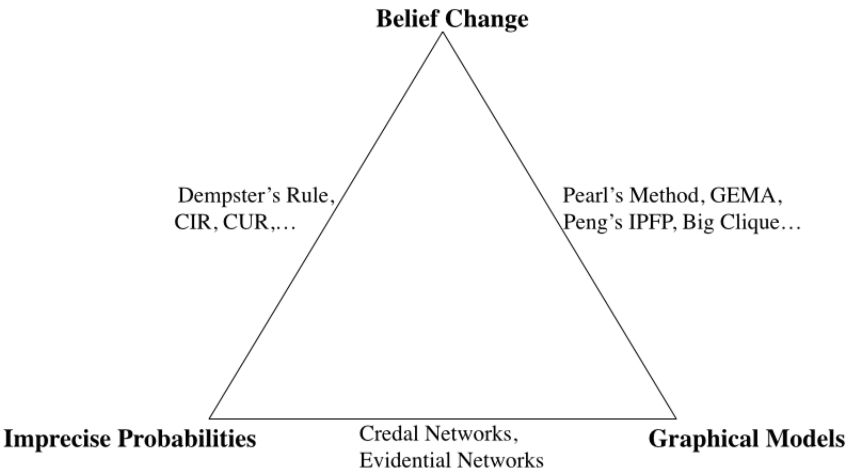
1. (Generalized) Belief Change: asymmetric adjustment of a belief base upon newly available information. We first considered the static approach to belief change (Belief Revision), and successively extended our interest to the cases of i) what we defined as *Partiality-inconsistent* evidence (Generalized Imaging), ii) overlapping findings on the same element of an agent's system of knowledge (Opinion Pooling), and, finally, iii) non-overlapping multiple observations (Iterated and Multiple Belief revision);

2. Graphical models: multivariate statistical models for probabilistic (counterfactual) reasoning. The key feature of such models identifies with the computational process being made efficient by the modular treatment of the complex, high-dimensional set of random variables, based on the pattern of conditional independence relationships among pairs of elements;

3. Imprecise Probabilities: based on the seminal work of Walley [257], they constitute a generalized approach to probabilistic reasoning, defined by sets of linear constraints. These in turn yield coherent collections of consistent probability mass functions over a (possibly joint) sample space.

A vast literature is available on imprecise treatment of probabilistic graphical models, originating from the seminal paper on Credal Networks of Cozman [51]. Also, Evidential networks were proposed, based on belief functions; see, among others [266, 268, 267]. Furthermore, several works were aimed to bridge belief change theory to graphical models, for a generalized treatment of evidence in the latter; among others, results from [140, 250, 244, 197] were repeatedly cited throughout Ch. 3 and

5. Finally, probabilistic belief change with imprecise probabilities was the subject of a wide range of contributions, including [224, 165, 276, 272, 78, 271] on belief revision, [210, 123] on partiality-inconsistent evidence, [2], and related works, on belief merging (or pooling). This is summarized by the figure below:



Although this is not the first work addressing a unified treatment of the three subjects listed above, (to our knowledge) previous works were all based on belief functions [32, 230]. These latter were proved to be outperformed by credal sets, that serve as a a more powerful and general tool for modeling imprecision [7].

To summarize, in our work, we first introduced basic concepts from the theory of sharp and imprecise probabilities. A major light was shed on the different concepts of irrelevance that go alongside with each of these, and their fulfillment of graphoid axioms, as a pre-requisite for graphical (DAG-based) modeling. This was successively introduced, and complexity of inference tasks with Bayesian and Credal networks was considered. An axiomatic introduction to belief change was the subject of the first part of Ch. 2, where adjustment operators, or rules, were characterized based on their kinematical properties, as opposed to a merely distance-based approach. Formal definition of soft evidence and of its extension to the conditional and imprecise case were provided, each accompanied by desirable kinematical properties for belief change. As a first contribution, we proposed the so-called *Imaging* operators for belief change, providing the counterpart of both sharp and imprecise Jeffrey's rule and Adams conditioning, with, once again, a focus on their kinematical properties. Implementation of standard and *credal* Jeffrey's rule, Adams conditioning was proposed throughout Ch. 3, based on the key result of Chan and Darwiche [35] on the inter-reducibility between soft and virtual evidence. This latter accounts for unreliability in the observational process. Treatment of virtual evidence and soundness of its propagation via message-passing procedure through a Bayesian network were introduces and established, respectively, by Pearl [194]. We proposed a number of transformations to perform propagation of uncertain evidence (both virtual and soft) in the generalized settings of conditional and credal findings, based on a sharp or imprecise graphical model. The discussion was successively specialized to the case of opinion pooling, where our methodology was proved to simultaneously i) combining overlapping instances, and ii) propagating them through a Bayesian network, while satisfying desirable kinematical properties. Finally, we moved from the static setting to the case of several findings being provided. We distinguished iterated from multiple belief revision, when evidence was provided together with an ordering, assigning a priority to each uncertain observation. We outlined kinematical principles that ought to be considered in such a generalized framework, and provided results on generalized forms of evidence that make the process *commutative*, i.e. that makes iterated procedures boil down to multiple belief revision. Finally, results on the complexity of inference were provided for the special case of multiple findings on mutually independent elements of the network.

We hope our work might contribute in opening the way to a number of research topics, that will be of great interest for future developments, both under a methodological and applied perspective. Some remarks and considerations are outlined in this merit at the end of every chapter. Among others, implementation of Imaginary operators for belief revision to graphical models, let them be Bayesian or Credal networks would yield to *closing the loop* upon the three vertices from the Figure above. The effort would likely provide useful tools for robust support to statistical decision making, record linkage, imputation, and so forth. Also, propagation of uncertain beliefs in a Credal network will likely be addressed in our future research, both under a theoretical point of view (with exact and approximate approaches), and to real-world problems, involving those settings affected by uncertainty and imprecision, cumulated from different sources. Furthermore, performance comparisons of our proposals with non-kinematical approaches might result of great interest, particularly when it comes to opinion merging and non-static frameworks, i.e. the iterated and multiple cases. This would not only include accounting for distance-based approaches, based on projections, but also tapping into the literature of belief fusion (that shall be intended as the symmetric extension of our approach). Finally, results on generalized forms of independence from Ch. 5 shall be extended to the imprecise setting. On a higher level, non-static iterated and multiple belief revision are expected to constitute a key topic for future research, with imprecise as well as sharp probabilities, involving a deep understanding of the mechanics that ought to be considered as benchmark. In this account, we proposed relaxing the responsiveness principle, i.e. retainment of acquired evidence, in favor of rigidity.

Appendix: Applications

1 An Application of Latin Hypercube Sampling for Learning Parameters of a Complex Model under Uncertainty: Modeling HPV

Latin Hypercube Sampling (LHS) is a methodology that belongs to the Monte Carlo class of procedures for the propagation of uncertainty. It was first introduced by Mckay in [180], extending *quota sampling* [236] as well as Latin square sampling [205]. As LHS displays desirable features and properties from both simple random and stratified sampling, it serves as a reasonable compromise between the two [129]. Briefly, let K be any (large) number of unknown parameters to estimate. LHS randomly generates N input vectors from the K-dimensional space of parameters, to generate a so-called *LHS matrix*. For each probability distribution D_i, and fixed N, LHS first generates a partition of the range of variation of D_i into N (mutually disjoint) equal probability intervals: $\Omega_{X_i} = \cup_{j=1}^{N} \delta_{i,j}$, $\delta_{i,j} \cap \delta_{i,j'} = \emptyset$, for every $j \neq j'$, $j = 1, \ldots, N$, $i = 0, \ldots, n$. Then, a single value is sampled from each $\delta_{i,j}$ without replacement, so that the whole range of variation of D_i is represented N values across the ordered partition; see [137] on a different approach.

Example 21. *Let $D_i \equiv P_{X_i}$ be a uniform distribution for random variable X_i over a continuous possibility space, $\Omega_{X_i} = [l_{X_i}, u_{X_i}]$, $-\infty < l_{X_i} < u_{X_i} < \infty$. Let $l_{X_i} = 0.0, u_{X_i} = 3.5$, and $N = 7$, $\Omega_{X_i} = [0.0, 3.5]$ is partitioned into:*

$$[0.0, 0.5) \cup [0.5, 1.0) \cup [1.0, 1.5) \cup [1.5, 2.0) \cup [2.0, 2.5) \cup [2.5, 3.0) \cup [3.0, 3.5]$$

By LHS, a single value is sampled uniformly from each interval, and the i-th column of the LHS matrix results as a permutation of

$$[0.013, 0.897, 1.152, 1.610, 2.406, 2.512, 3.078]'$$

The N values sampled for each $X_i \in \mathbf{V}$ are combined at random, and constitute the rows of the LHS matrix in $\mathbb{R}^{N \times K}$. The model is run N times, one for each configuration (row of the matrix). Let $\hat{\phi}$ be the estimate obtained by LHS, its variance was proved to be always smaller compared to any other obtained by stratified sampling [181], which in turn outperforms applications of simple random sampling; this motivates LHS as a more accurate, while efficient, technique, compared to the other two considered. When uncertainty on the K parameters induce bounds that are considered too loose, or imprecise, LHS may be iterated to reduce their size: the N outputs evaluated based on a properly chosen functional, and the parameter sets associated with the *best performances* are used to refine the ranges of variation of each quantity. This procedure is iteratively applied until some convergence is reached.

We resort to LHS to identify a subset of $K = 14$ parameters of a compartmental model.

Deterministic compartmental models were introduced in epidemiology by Kermack and McKendrick [149], to describe the spread of an infectious disease in a wholly susceptible (fixed) population. In the simplest case, such model are described by systems of ODEs. Consider the SIR flow-diagram, in Fig. 5.3. In this model individuals can move through a discrete sequence of states - or compartments - namely the susceptible compartment, the infective compartment etc. To each compartment we associate a states variables S(t),I(t) etc representing the numbers of individuals who at time t are susceptible to the infection, infective, and recovered, respectively. With time, people move from a compartment to another according to some appropriate rate; e.g., the dynamics of the SIR model from Fig. 5.3 are described by the following:

$$\frac{\partial S(t)}{\partial t} = -\lambda S(t)$$
$$\frac{\partial I(t)}{\partial t} = \lambda S(t) - \gamma I(t)$$
$$\frac{\partial R(t)}{\partial t} = \gamma I(t)$$

It is straightforward to see the total population is constant over time: $\frac{\partial S(t)}{\partial t} + \frac{\partial I(t)}{\partial t} + \frac{\partial R(t)}{\partial t} = 0$, for all $t \geq 0$.

λ is called *force of infection*; it corresponds to the rate at which any susceptible individual acquires a given disease following a contact with an infectious individual[21]. It is a function of the number of individuals in compartment I at each time step, of the number of contacts and of the infectiousness of the disease itself per contact[22]. Let $N_j(t)$ be the overall number of individuals in compartment j, at some point t

[21] The definition of contact changes with the disease, and the way it propagates through a community.

[22] See [6, 18] for details.

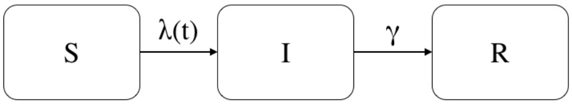


Figure 5.3: The SIR model without demographic dynamics.

in time,

$$N_j(t) \sim \ \text{Poisson}(\delta(t))$$

where

$$\delta(t) = N_j(0)e^{\theta t} \tag{5.10}$$

with $N_j(0)$ and θ being, respectively, the size of compartment j at time $t = 0$, and the associated growth rate.[23] Eq. (5.10) may be replaced with other functionals than the exponential linear growth, such as a quadratic growth, a Weibull model, a logistic growth or a Gompertz growth, to properly fit available data.

Roughly, each individual in a compartmental model takes a random walk over a Markov chain, as a consequence of the *memoryless* property of all Poisson processes. At each time step, the overall model represents the population, partitioned into mutually exclusive homogenous groups; e.g., in the SIR model, $S(t) + I(t) + R(t) = P$, for each $t \geq 0$. See [131, Ch.3] for a soft introduction to mathematical modeling of infectious diseases.

Epidemic compartmental models may be specialized to describe infectious disease whose etiology and epidemiology are more complex than the SIR of Fig. 5.3, i.e., susceptible to infectious to (permanently) removed. To start with, the population may be stratified according to age classes, gender, behavior, and demographic dynamics such as births and deaths (natural and, eventually, caused by the disease), migration flows ought to be included. Also, further compartments may be introduced, to properly describe the natural history of an infectious disease; e.g. newborns may be immune to a disease, thanks to maternal antibodies and/or the disease itself may be characterized by different stages, whose infectiousness, dynamics, morbidity or mortality ought to be separately considered.

Compartmental models are used for a wide range of purposes. Among others, they are used by policy makers to evaluate the impact, through time and at the population level, of the introduction of (one or more) control measures, e.g., vaccination, screening. A model is initiated by introducing an infectious individual[24] in a fully

[23] $N_j(t)$ may be equivalently specified as a generalized linear model with logarithmic link function and Poisson distribution for the response.

[24] In the general case, more than an infectious individual may be introduced, based on features

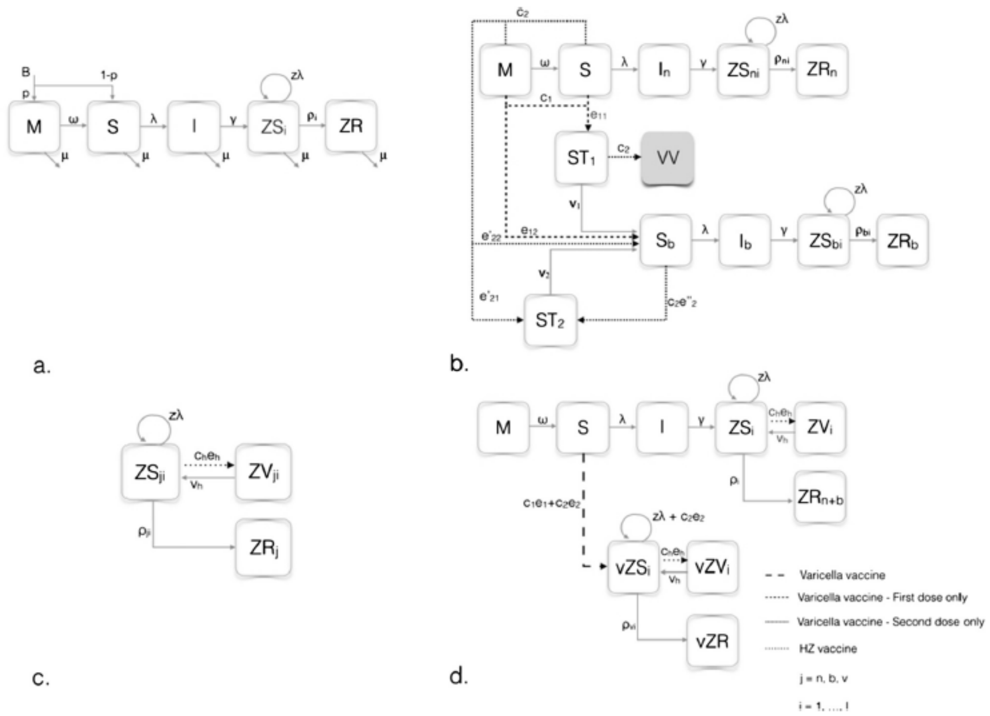


Figure 5.4: Flow-diagram of the model used in [173]. a) Extension of the standard SIR model with demographic dynamics; b) Dynamics induced by the introduction of the anti-varicella vaccine; c) Dynamics induced by the anti-Herpes Zoster vaccination; and d) Dynamics related to the varicella cases from vaccine strain.

susceptible population. The individual induces secondary cases, according to a key quantity denoted as R_0, that in turn increasingly propagate the disease through the population until a so-called *endemic equilibrium* is reached [6]. There, if available, control measures may be incorporated in the model, by means properly specified additional compartments. See Fig. 5.4 from [173] as an example, where it is modeled the spread and behavior of the Varicella Zoster Virus (VZV) in a given population. Panel a), particularly, represents an extension of the standard SIR model with demographic dynamics, where individuals enter a sequence of further stages of susceptibility to Herpes Zoster (ZS_i) before they eventually acquire full immunity to the natural VZV-associated diseases. Panel b) illustrates the dynamics induced by the introduction of the anti-varicella vaccine, whereas panel c) illustrates the anti-Herpes Zoster vaccination. Finally, panel d) illustrates the dynamics related to the occurrence of diseases from vaccine strain.

A major issue with compartmental ODE-based models is their identification, i.e., full specification of the parameters regulating the random walks represented by the

such as the infectiousness of the disease and the contact pattern that characterize the population, among others.

sequences of compartments. Usually, the structure of a model results from experts' review on the etiology and epidemiology of the disease. Then, if available, disease-related quantities are derived from the literature, e.g., pattern of (potentially infectious) contacts of the population, infectiousness of the disease and time spent as infectious, as well as efficacy and waning rate of a vaccine. Still, it is not unlikely that a (possibly very large) number of parameters is unknown and needs to be estimated.

Our application considers an ODE-based compartmental model, built to reproduce the heterosexual transmission dynamics of *human papilloma virus* (HPV) infections caused by 9 genotypes, and progression to cervical cancer from infection through various stages of disease in a stationary population. HPV is a DNA virus whose infections may result in precancerous and cancerous lesions. HPV is the most common sexually transmitted infectious disease [182], although most infections follow a course that naturally resolves. Those that proceed to lesions expose infected women to an increased risk of cervical cancer. Among the so-called *high-risk* genotypes, HPV-16 and 18 were found to be responsible alone for almost 80% of all cervical cancers worldwide [122]. We consider the population stratified by gender, 3 sexual activity levels and 100 unitary age classes. As a consequence, any compartment needs to be specified for each of the hundreds possible combinations, yielding overparametrization of the model. e.g., let a, g, l denote, respectively, age, gender and sexual activity level, $S(a, g, l, t)$ needs to be specified for each combination (a, g, l), at every $t \geq 0$.[25]

We considered the ODE-based compartmental model of [124], to evaluate the impact of a nonavalent vaccine targeting most of high-risk genotypes on the population from Italian region Puglia (Apulia). We compared different prevention strategies, such as a bivalent vaccine [124] and/or a combined ongoing screening campaign.

The model, whose simplified flow-diagram is reported in Fig. 5.5 for fixed age a and sexual activity level l, describes a complex pattern of dynamics, and required detailed parametrization. Although most values were specified based on the literature[26], a large number ($K = 14$) of parameters was left free for estimation. Particularly, $K_1 = 6$ referred to the transmission process of HPV, whereas $K_2 = 8$ served to calibrate the progression/regression dynamics across different stages for the lesions (CIN1 to CIN2 to CIN3, acronyms for Cervical Intraepithelial Neoplasia, to cancer *in situ*, CIS, or cervical, CC). We considered data on i) prevalence of HPV, ii) reported incidence of cancer of the cervix, in the female component of the Apulian

[25] Some simplifications may be applied. For example, sexual activity level needs no specification for people aged 11 or less.

[26] Based on regional and national sources.

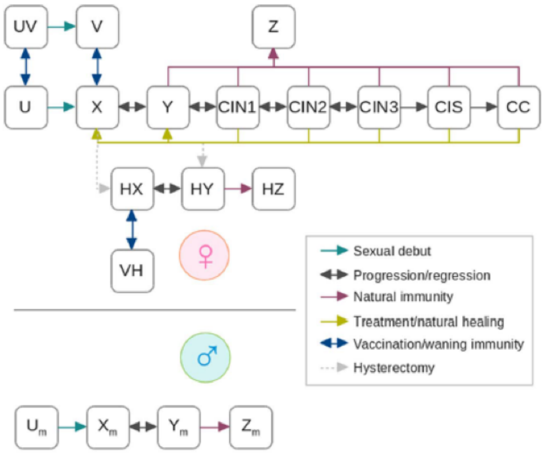


Figure 5.5: Simplified flow-diagram from [124] to model the spread and behavior of genotypes 16-18 of Human Papilloma Virus. The same dynamics were considered to model the genotypes involved by the nonavalent anti-HPV vaccine.

population.[27]
The estimation process was carried out based on iterated LHS. Following previous works of [174, 124], we considered two successive procedures. As a first step, K_1 parameters on the transmission dynamics, or *natural history parameters*, were estimated, and the (logarithmic) Poisson likelihood of the age-specific HPV prevalence predicted by the model for each row of the $N \times K_1$ LHS matrix was computed, with $N = 10,000$. Let $\hat{\mathbf{x}}, \mathbf{x} \in \mathbb{R}^{A+1}$ be the number of age-specific HPV cases predicted by the model and available as data[28], respectively:

$$\log(\hat{\mathbf{x}}; \mathbf{x}) \propto \sum_{a=0}^{A} [\hat{x}_a \cdot \log(x_a) - x_a]$$

Ranges of uncertainty associated with each parameter were initiated as vacuous, or very loose at least; e.g., values such as probabilities or proportions were considered as varying between 0 and 1. At each iteration, the best configurations (based on the empirical cumulative distribution function obtained) were used to refine the ranges of variation, until convergence (see Table 5.1). Second step of the estimation process was applied holding the K_1 estimated parameters on natural history fixed. This way, search in the K dimensional space of parameters was restricted in a K_2 dimensional region whose predictions were always consistent with the age-specific

[27]Data were obtained from a pooling and re-weighting process of datasets available from the literature that did not explicitly refer to the genotypes targeted by the nonavalent vaccine; see [172] for a thorough discussion on this application.
[28]$\hat{\mathbf{x}}$ was generated for all unitary age classes, although $\mathbf{x}$ was available for larger classes. The corresponding average value of $\hat{\mathbf{x}}$ was used.

Parameter	Estimated Value
Probability of Male-to-Female Transmission	94.1%
Probability of Female-to-Male Transmission	96.2%
Coefficient of Age Assortativity	0.682
Coefficient of Sexual Activity Level Assortativity	0.683
Probability to Acquire Natural Immunity	90.1%
Average Duration of Lesion-free Infections	2.66 yr
Progression Rate HPV-to-CIN1	0.007 yr−1
Progression Rate CIN1-to-CIN2	0.016 yr−1
Progression Rate CIN2-to-CIN3	0.004 yr−1
Progression Rate CIN3-to-CIS	0.237 yr−1
Progression Rate CIS-to-CC	0.007 yr−1
Regression Rate from CIN1	0.212 yr−1
Regression Rate from CIN2	0.600 yr−1
Regression Rate from CIN3	0.88 yr−1

Table 5.1: Parameter estimates obtained by the LHS procedure.

HPV prevalence curve observed. The second LHS matrix generated had K_2 columns. Estimates on the incidence of cancer (both CIS and CC) were compared with data based on the root mean squared relative error functional:

$$E_{1,2} = \sqrt{\frac{\sum_{a=0}^{A} \left(\frac{x_a - \hat{x}_a}{x_a}\right)^2 + \sum_{b=0}^{B} \left(\frac{x'_b - \hat{x}'_b}{x'_b}\right)^2}{A + B}}$$

with $\hat{\mathbf{x}}', \mathbf{x}'$ in $\mathbb{R}^B$ vectors of estimated and observed cases of cervical cancer over $B + 1$ age classes, as with prevalence.

The fully parametrized model was used to evaluate the impact of several combined control measures (including immunization by different vaccines and screening) on the incidence of cervical cancer in the female component of the population; see Fig. 5.6. Also, the model provided insights on changes in the age-specific incidence with time, as Fig. 5.7 shows. Both figures were taken from [172], where details may be found.

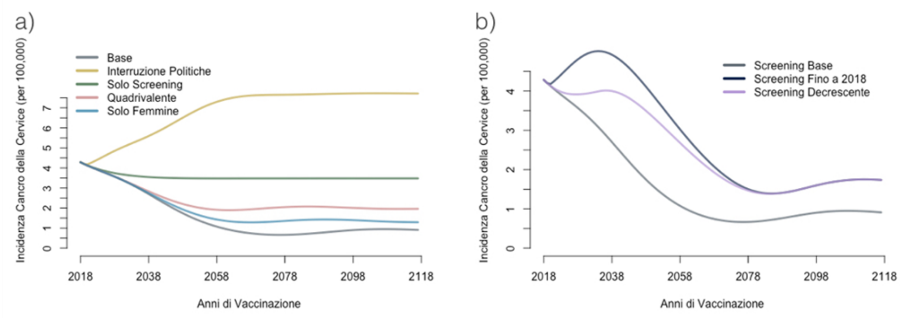


Figure 5.6: Average yearly incidence (per 100,000 women) of cervical cancer under different scenarios of control, from 2018 to 2117. Panel a): 1) baseline (purple), 2) shut-down of all control measures from 2017 (gold), 3) baseline screening only (green), 4) baseline screening combined with 4-valent vaccination, introduced in 2008(pink), 5) baseline vaccination restricted to the sole female component of the population (light blue). Panel b) baseline vaccination combined with baseline (grey), decreasing in 15 years (purple) and shut-down (blue) screening.

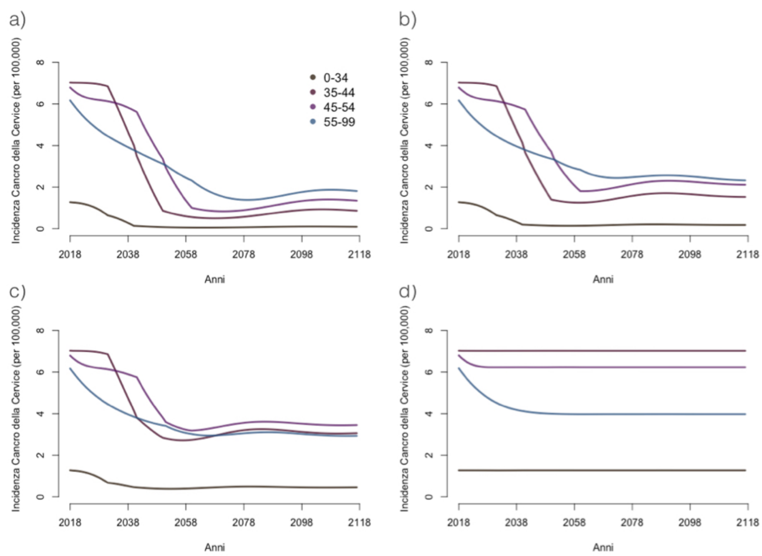


Figure 5.7: Yearly incidence of cervical cancer (per 100,000 women) for age classes 0-34, 35-44, 45-54 e 55-99, according to the baseline scenario (panel a), baseline vaccination scenario restricted to the sole feminine component of the population (panel b), 4-valent vaccination (ongoing since 2008, panel c), screening only from 2018 onwards (panel d).

2 An Application of Context-Specific Independence Relationships Detection: CSeek

D-separation is a sound and complete method for detecting conditional SI in a BN. We hereby briefly introduce CSeek, a routine for the detection of local CSI relationships within a BN, extending previous work from [95] to acyclic directed graphical structures.
The procedure is intended to be applied within the structural learning process of the DAG by the Conservative PC algorithm (CPC) [209], although it may be readily applied on any available structure, either fully or partially directed. Roughly, the CPC algorithm is a constraint-based structural learning process that tackles failure of faithfulness condition. It extends the well-known PC algorithm [234] by marking *ambiguous* unshielded triples as *unfaithful*, whenever faithfulness is likely to be failed by PMF P. As a consequence, CPC produces a graph whose independence pattern is consistent with several Markov equivalence classes [234], rather than PC's unique partially oriented DAG[29]; see [234, 209] for details. To keep the discussion as general as possible, suppose every variable X is associated Markov blanket $mb(X)$, whose elements are parents, children (and potential parents of X, if the graph is partially oriented). CSeek is a routine that compiles a list of CSI relationships; and thus naturally fits within the framework of structural learning of LDAGs (see Sec: 1.3).
We hereby provide a synthetic representation of oracle CSeek: statistical tests ought to be included within the **if**-step of the routine below, given a sample.

Definition 54 (CSeek)**.** *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be any DAG. Without loss of generality, suppose the graph is partially directed*[30]*. Each node X is associated a set of adjacent nodes $Ch(X) \cup (Pa(X) \cup Adj(X)) = Ch(X) \cup Pa^+(X) = mb^+(X)$. Let $\mathcal{L}_\mathbf{E}$ be the (empty) list of labels that are to be attached to edge/arc $(Y, X) \in \mathbf{E}$ whenever it holds $I(X, Y; \mathbf{W} = \mathbf{x}^*)$, for some $\mathbf{w}^* \in \Omega_\mathbf{W}$, with $\{Y, \mathbf{W}\} \subseteq Pa^+(X)$.*

$\mathcal{L}_\mathbf{E} := \emptyset$; $\mathbf{D} := \emptyset$
for all $i = 0, ..., n$ ***do***
  $mb^+(X_i) := mb^+(X_i) \backslash Ch(X_i)$
  $\mathbf{D}_X := \Omega_{mb^+(X)}$
  $s := 0$
  while $|mb^+(X)| \geq s + 2$ ***do***

[29] A partially oriented DAG is a graphical structure used to identify a *Markov equivalence class* [234].
[30] If CSeek fits as an additional step to CPC, *unfaithful triples* need un-marking prior to application of the routine.

$\mathbf{D}_{XY} := \Omega_{mb^{+}(X)\setminus\{Y\}}$
$s := s + 1$
for all $\mathbf{d}_{XY} \in \mathbf{D}_{XY}$ ***do***
  for all $\mathbf{W} \in mb^{+}(X)\setminus\{Y\} : |\mathbf{W}| = s$ ***do***
    if $I(X, Y|\mathbf{W} = \mathbf{w}), \mathbf{w} \sim \mathbf{d}_{XY}$ ***then***
      $\mathcal{L}_{Y,X} := mb^{+}(X)\setminus\{Y, \mathbf{W}\} \cup \{\mathbf{W} = \mathbf{w}\}$
      $\mathcal{L}_{\mathbf{E}} := \mathcal{L}_{\mathbf{E}} \cup \mathcal{L}_{Y,X}$
      $\mathbf{D}_{XY} := \mathbf{D}_{XY}[\mathbf{W} \not\sim \mathbf{w}]$
      $Ch(Y) := Ch(Y) \cup \{X\}$
      $Ch(\mathbf{W}) := Ch(\mathbf{W}) \cup \{X\}$
    end if
  end for
end for
end while
end for

Step 3 of the CPC algorithm follows.
Under proper consistency conditions, the integrated procedure results in a Completed Partially Labeled DAG (CPLDAG), inducing a *CSI Equivalence Class* [11].

Example 22. *Consider the example from [209]. We are given records on the occurrences of thrombosis (T) among women, accounting for their behavior relevant to the disease: whether they take birth control pills (B) and they are pregnant (P); see Fig. 5.8 (left panel). It is known that both taking birth control pills and being pregnant increase the chances to develop the disease. This results in the paths going from node B to P canceling out.*
As a consequence, the structure is not faithful to the true probability distribution, as the latter encodes CI $I(T, B)$. In the general case, the CPC algorithm, as well as the PC algorithm, learns the clique of Fig. 5.8 (middle panel) from data, and thus fails to identify the marginal independence between node B and T.[31]
We apply the CPC algorithm integrated with CSeek, provided ordering $\{T, B, P\}$. After the first iteration, the algorithm stops: label $\mathcal{L}_{BT} = \{P = true\}$ results and arcs (B, T) and (P, T) are oriented accordingly as parents of node T, as in Fig. 5.8. The (partially oriented) LDAG resulting provides a pattern of independence which is closer to the true probability distribution than a clique, as it encodes CSI

[31] It is worth noting how learning any structure other than the clique would have yielded no better results, as it would have borne additional unwanted CI relationships. Also, directions are dropped as any orientation of the arcs, although inducing a v-structure while avoiding cycles, would fall within the same Markov equivalence class.

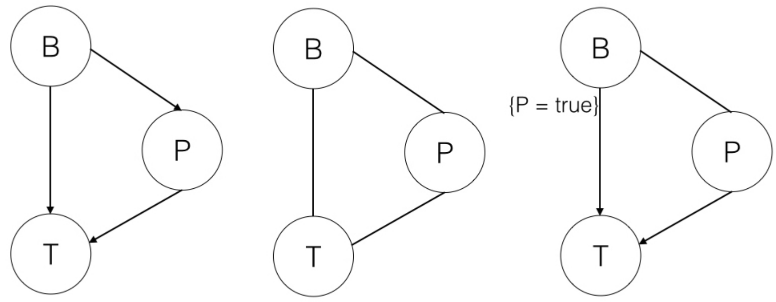


Figure 5.8: Toy example from [209]. Binary variables B, P and T represent, respectively, Birth Control Pill, Pregnancy and Thrombosis, with $\Omega_j = \{\text{true}, \text{false}\}$, $j = B, P, T$. The structure is unfaithful to the true probability distribution as P encodes $I(B, T)$. Left panel: The true DAG; Middle panel: The clique resulting from application of standard CPC algorithm; Right panel: The LDAG resulting from CPC integrated with CSeek.

$I(T, B|P = true)$. *The PMF fitted on the structure learned by standard CPC has* $2^3 - 1$ *free parameters, whereas the one accounting for CSIs has* $2^2 + 1$.

Elicitation of local CSIs allows for a parsimonious representation of the dependency model. As our routine was developed within the framework of the constraint based structure learning, it comes naturally in the style of the PC algorithm. Nonetheless, even under faithfulness condition, sparsity of the graph to be learned is critical for consistency and computational feasibility of the PC algorithm [145] and thus it is likely to be outperformed by heuristic strategies under fairly general conditions. In order not to suffer from this sort of limitations, we explored the performance of our procedure when applied to a fully defined known DAG.

Example 23. *We report in Table 5.2 results on benchmark BNs from the literature. In our experiments, data records were generated from each compiled network, and sample local CSIs were elicited.*
Application of CSeek allows for a more parsimonious representation of the network and thus requires a significantly reduced number of computations to be performed when answering a probabilistic query on the elements of the model. As threshold parameter ϵ is increased, robustness of the labeling procedure grows as well, since we label as context-specific independent only those pairs of variables whose associated edge is (significantly) weak under specific instantiations. Conversely, labeling those edges that are only slightly weak might result in a too strong approximation and thus worsen the performance of the inferential tasks, e.g., classification.

Detection of local CSIs when learning a network might result in a constrained

Network	$\|V\|$	k	No. Distinct Parameters	$\epsilon = 0.01$	$\epsilon = 0.1$
Sachs	11	3	178	66 (-62.9%)	86 (-51.7%)
Alarm	37	4	509	370 (-27.3%)	396 (-22.2%)
Hepar	70	6	1453	695 (-52.2%)	779 (-46.4%)

Table 5.2: Reduced parametric dimensionality in benchmark networks, with thresholds $\epsilon = 0.01$ and $\epsilon = 0.1$ for G^2 deviance statistic for conditional SI. Complete records were simulated from the networks, with fixed sample size $M = 1000$.

parametrization of the model [191], significantly reducing the complexity of the inferential tasks. Also, it allows for approximate inferences, particularly when CSIs are elicited from data and the network is already known, as in, e.g., [199]. CSeek constitutes a simple tool for eliciting knowledge on context-specific behavior from data, that may be readily exploited within our framework of probabilistic belief revision of a PGM.

Bibliography

[1] Resolving peer disagreements through imprecise probabilities. *Noûs*, 52(2):260–278, 2018.

[2] M. Adamčik and G. Wilmers. The irrelevant information principle for collective probabilistic reasoning. *Kybernetika*, 50(2):175–188, 2014.

[3] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985.

[4] Russel G Almond. *Graphical belief modeling.* CRC Press, 1995.

[5] JD Alves, JF Ferreira, J Lobo, and J Dias. Brief survey on computational solutions for bayesian inference. In *Workshop on Unconventional computing for Bayesian inference*, 2015.

[6] Roy M Anderson, Robert M May, and B Anderson. *Infectious diseases of humans: dynamics and control*, volume 28. Wiley Online Library, 1992.

[7] Alessandro Antonucci. An interval-valued dissimilarity measure for belief functions based on credal semantics. In *Belief Functions*, pages 37–44. Springer, 2012.

[8] Alessandro Antonucci, Cassio P De Campos, David Huber, and Marco Zaffalon. Approximating credal network inferences by linear programming. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 13–24. Springer, 2013.

[9] Alessandro Antonucci, Yi Sun, Cassio P de Campos, and Marco Zaffalon. Generalized loopy 2u: a new algorithm for approximate inference in credal networks. *International Journal of Approximate Reasoning*, 51(5):474–484, 2010.

[10] Alessandro Antonucci and Marco Zaffalon. Equivalence between bayesian and credal nets on an updating problem. *Soft Methods for Integrated Uncertainty Modelling*, pages 223–230, 2006.

[11] Alessandro Antonucci and Marco Zaffalon. Decision-theoretic specification of credal networks: A unified language for uncertain modeling with sets of bayesian networks. *International Journal of Approximate Reasoning*, 49(2):345–361, 2008.

[12] Horacio Arló-Costa and Richmond H Thomason. Iterative probability kinematics. *Journal of Philosophical Logic*, 30(5):479–524, 2001.

[13] Kenneth J Arrow. A difficulty in the concept of social welfare. *Journal of political economy*, 58(4):328–346, 1950.

[14] Kenneth Joseph Arrow. *Social Choice and Individual Values.* Number 12. Yale University Press, 1963.

[15] Thomas Augustin, Frank PA Coolen, Gert de Cooman, and Matthias CM Troffaes. *Introduction to imprecise probabilities.* John Wiley & Sons, 2014.

[16] Michael Bacharach. Scientific disagreement. *Unpublished manuscript*, 1972.

[17] Michael Bacharach. Group decisions in the face of differences of opinion. *Management Science*, 22(2):182–191, 1975.

[18] Norman TJ Bailey et al. *The mathematical theory of infectious diseases and its applications.* Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.

[19] Jon Barwise. An introduction to first-order logic. In *Studies in Logic and the Foundations of Mathematics*, volume 90, pages 5–46. Elsevier, 1977.

[20] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Phil Trans R Soc*, 53:370–418, 1763.

[21] Alessio Benavoli and Alessandro Antonucci. Aggregating imprecise probabilistic knowledge. In *Proc. 6th International Symposium on Imprecise Probability*, 2009.

[22] Salem Benferhat, Didier Dubois, Souhila Kaci, and Henri Prade. Possibilistic merging and distance-based fusion of propositional information. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):217–252, 2002.

[23] Jean-Marc Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39(2-3):123–150, 2005.

[24] Lawrence Blume, Adam Brandenburger, and Eddie Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica: Journal of the Econometric Society*, pages 61–79, 1991.

[25] Craig Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25(3):263–305, 1996.

[26] Craig Boutilier, Nir Friedman, Moises Goldszmidt, and Daphne Koller. Context-specific independence in bayesian networks. In *Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence*, pages 115–123. Morgan Kaufmann Publishers Inc., 1996.

[27] Richard Bradley. Radical probabilism and Bayesian conditioning. *Philosophy of Science*, 72(2):342–364, 2005.

[28] Richard Bradley, Franz Dietrich, and Christian List. Aggregating causal judgments. *Philosophy of Science*, 81(4):491–515, 2014.

[29] Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

[30] Andrés Cano, Manuel Gómez, Serafín Moral, and Joaquín Abellán. Hill-climbing and branch-and-bound algorithms for exact and approximate inference in credal networks. *International Journal of Approximate Reasoning*, 44(3):261–280, 2007.

[31] Andrés Cano and Serafín Moral. A review of propagation algorithms for imprecise probabilities. 1999.

[32] José Cano, Miguel Delgado, and Serafin Moral. An axiomatic framework for propagating uncertainty in directed acyclic networks. *International Journal of Approximate Reasoning*, 8(4):253–280, 1993.

[33] Andrea Capotorti, Giuliana Regoli, and Francesca Vattari. On the use of a new discrepancy measure to correct incoherent assessments and to aggregate conflicting opinions based on imprecise conditional probabilities. In *Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications, ISIPTA*, 2009.

[34] Hei Chan and Adnan Darwiche. A distance measure for bounding probabilistic belief change. *International Journal of Approximate Reasoning*, 38(2):149–174, 2005.

[35] Hei Chan and Adnan Darwiche. On the revision of probabilistic beliefs using uncertain evidence. *Artificial Intelligence*, 163(1):67–90, 2005.

[36] Jake Chandler. Subjective probabilities need not be sharp. *Erkenntnis*, 79(6):1273–1286, 2014.

[37] Pengwen Chen, Yunmei Chen, Murali Rao, et al. Metrics defined by bregman divergences. *Communications in Mathematical Sciences*, 6(4):915–926, 2008.

[38] Jian Cheng and Marek J Druzdzel. Latin hypercube sampling in bayesian networks. In *FLAIRS Conference*, pages 287–292, 2000.

[39] Arthur Choi, Khaled S Refaat, and Adnan Darwiche. Edml: A method for learning parameters in bayesian networks. *arXiv preprint arXiv:1202.3709*, 2012.

[40] J. Cleland. *Orthopaedic Clinical Examination: An Evidence-Based Approach for Physical Therapists*. Elsevier, 2005.

[41] Flávio Codeço Coelho and Claudia T Codeço. Dynamic modeling of vaccinating behavior as a function of individual beliefs. *PLoS computational biology*, 5(7):e1000425, 2009.

[42] Giulianella Coletti and Romano Scozzafava. Zero probabilities in stochastic independence. *Information, Uncertainty, Fusion*, pages 185–196, 2000.

[43] Giulianella Coletti and Romano Scozzafava. Exploiting zero probabilities. In *Probabilistic Logic in a Coherent Setting*, pages 117–126. Springer, 2002.

[44] Patrick L Combettes. The foundations of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208, 1993.

[45] G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.

[46] Gregory Floyd Cooper. *Bayesian belief-network inference using recursive decomposition.* Knowledge Systems Laboratory, Medical Computer Science, Stanford University, 1992.

[47] G Corani, A Antonucci, and M Zaffalon. Bayesian networks with imprecise probabilities: Theory and application to classification. *Data Mining: Foundations and Intelligent Paradigms*, 23:49–93, 2012.

[48] Inés Couso and Serafín Moral. Sets of desirable gambles: conditioning, representation, and precise probabilities. *International Journal of Approximate Reasoning*, 52(7):1034–1055, 2011.

[49] Inés Couso, Serafín Moral, and Peter Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5(2):165–181, 2000.

[50] Mikaël Cozic. Imaging and sleeping beauty: a case for double-halfers. *International Journal of Approximate Reasoning*, 52(2):137–143, 2011.

[51] Fabio G. Cozman. Credal networks. *Artificial Intelligence*, 120(2):199–233, 2000.

[52] Fabio G Cozman. Separation properties of sets of probability measures. In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, pages 107–114. Morgan Kaufmann Publishers Inc., 2000.

[53] Fabio G Cozman. Sets of probability distributions, independence, and convexity. *Synthese*, 186(2):577–600, 2012.

[54] Fabio G Cozman and Peter Walley. Graphoid properties of epistemic irrelevance and independence. *Annals of Mathematics and Artificial Intelligence*, 45(1):173–195, 2005.

[55] Fabio Gagliardi Cozman. Algorithms for conditioning on events of zero lower probability. In *FLAIRS Conference*, pages 248–252, 2002.

[56] Fabio Gagliardi Cozman and Teddy Seidenfeld. *Independence for full conditional measures, graphoids and Bayesian networks.* EPUSP, 2007.

[57] Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pages 146–158, 1975.

[58] J.C.F. da Rocha, A.M. Guimaraes, and C.P. de Campos. Dealing with soft evidence in credal networks. In *Proceedings of* Conferencia Latino-Americana de Informatica, 2008.

[59] José Carlos Ferreira da Rocha and Fabio Gagliardi Cozman. Inference in credal networks: branch-and-bound methods and the a/r+ algorithm. *International Journal of Approximate Reasoning*, 39(2-3):279–296, 2005.

[60] Jose Carlos Ferreira da Rocha, Fabio Gagliardi Cozmanl, and Cassio Polpo de Campos. Inference in polytrees with sets of probabilities. In *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*, pages 217–224. Morgan Kaufmann Publishers Inc., 2002.

[61] Paul Dagum and Michael Luby. Approximating probabilistic inference in bayesian belief networks is np-hard. *Artificial intelligence*, 60(1):141–153, 1993.

[62] Bruce D'Ambrosio. Symbolic probabilistic inference in large bn20 networks. In *Uncertainty Proceedings 1994*, pages 128–135. Elsevier, 1994.

[63] Adnan Darwiche. A differential approach to inference in bayesian networks. *Journal of the ACM (JACM)*, 50(3):280–305, 2003.

[64] Adnan Darwiche. *Modeling and reasoning with Bayesian networks.* Cambridge University Press, 2009.

[65] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. *Artificial intelligence*, 89(1-2):1–29, 1997.

[66] A Philip Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–31, 1979.

[67] A Philip Dawid. Conditional independence. *Encyclopedia of statistical sciences, update*, 2:146–153, 1998.

[68] AP Dawid, RG Cowell, SL Lauritzen, and DJ Spiegelhalter. *Probabilistic networks and expert systems*. Springer-Verlag, 1999.

[69] Cassio Polpo de Campos and Fabio Gagliardi Cozman. Inference in credal networks using multilinear programming. In *Proceedings of the Second Starting AI Researcher Symposium*, pages 50–61, 2004.

[70] Cassio Polpo De Campos and Fabio Gagliardi Cozman. The inferential complexity of bayesian and credal networks. In *IJCAI*, volume 5, pages 1313–1318, 2005.

[71] Cassio Polpo de Campos and Fabio Gagliardi Cozman. Computing lower and upper expectations under epistemic independence. *International Journal of Approximate Reasoning*, 44(3):244–260, 2007.

[72] Cassio Polpo de Campos and Fabio Gagliardi Cozman. Inference in credal networks through integer programming. In *Proceedings of the 5th International Symposium on Imprecise Probability: Theories and Applications*, pages 145–154, 2007.

[73] Luis M De Campos and Serafin Moral. Independence concepts for convex sets of probabilities. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 108–115. Morgan Kaufmann Publishers Inc., 1995.

[74] Luiz Max de Carvalho, Daniel AM Villela, Flavio Codeco Coelho, and Leonardo Soares Bastos. Choosing the weights for the logarithmic pooling of probability distributions. *arXiv preprint arXiv:1502.04206*, 2015.

[75] G. De Cooman and M. Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159(1-2):75–125, 2004.

[76] Gert De Cooman, Filip Hermans, Alessandro Antonucci, and Marco Zaffalon. Epistemic irrelevance in credal nets: the case of imprecise markov trees. *International Journal of Approximate Reasoning*, 51(9):1029–1052, 2010.

[77] Gert De Cooman, Enrique Miranda, and Marco Zaffalon. Independent natural extension. *Artificial Intelligence*, 175(12-13):1911–1950, 2011.

[78] Gert De Cooman and Marco Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159(1-2):75–125, 2004.

[79] Bruno De Finetti. La prévision: ses lois logiques, ses sources subjectives. In *Annales de l'institut Henri Poincaré*, volume 7, pages 1–68, 1937.

[80] Abraham De Moivre. *The Doctrine of chances: or, A Method of Calculating the probability of events in Play.* W. Pearson, 1718.

[81] Rina Dechter. Bucket elimination: A unifying framework for probabilistic inference. In *Learning in graphical models*, pages 75–104. Springer, 1998.

[82] Arthur P Dempster. Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics*, pages 325–339, 1967.

[83] M Di Zio, G Sacco, M Scanu, and P Vicard. Some approaches in imputing missing items with bayesian networks. *Atti della XLII Riunione Scientifica della Società Italiana di Statistica*, 2004.

[84] Persi Diaconis and Sandy L. Zabell. Updating subjective probability. *Journal of the American Statistical Association*, 77(380):822–830, 1982.

[85] Franz Dietrich. Judgment aggregation and agenda manipulation. *Games and Economic Behavior*, 95:113–136, 2016.

[86] Franz Dietrich. A theory of bayesian groups. 2016.

[87] Franz Dietrich and Christian List. Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 29(1):19–33, 2007.

[88] Franz Dietrich and Christian List. Probabilistic opinion pooling generalized part one: General agendas. 2013.

[89] Franz Dietrich, Christian List, and Richard Bradley. Belief revision generalized: A joint characterization of bayes' and jeffrey's rules. *Journal of Economic Theory*, 162:352–371, 2016.

[90] Igor Douven and Jan-Willem Romeijn. A new resolution of the Judy Benjamin problem. *Mind*, 120(479):637–670, 2011.

[91] Lester E Dubins. Finitely additive conditional probabilities, conglomerability and disintegrations. *The Annals of Probability*, pages 89–99, 1975.

[92] Didier Dubois. Three scenarios for the revision of epistemic states. *Journal of Logic and Computation*, 18(5):721–738, 2008.

[93] Didier Dubois and Henri Prade. Focusing vs. belief revision: A fundamental distinction when dealing with generic knowledge. In *Qualitative and quantitative practical reasoning*, pages 96–107. Springer, 1997.

[94] Kenny Easwaran. Bayesianism i: Introduction and arguments in favor. *Philosophy Compass*, 6(5):312–320, 2011.

[95] Alejandro Edera, Federico Schluter, and Facundo Bromberg. Learning markov networks with context-specific independences. In *Tools with Artificial Intelligence (ICTAI), 2013 IEEE 25th International Conference on*, pages 553–560. IEEE, 2013.

[96] Gal Elidan. Copula bayesian networks. In *Advances in neural information processing systems*, pages 559–567, 2010.

[97] Paul Embrechts, Er Mcneil, and Daniel Straumann. Correlation: pitfalls and alternatives. In *Risk Magazine*. Citeseer, 1999.

[98] Edward S Epstein. A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, 8(6):985–987, 1969.

[99] Enrico Fagiuoli and Marco Zaffalon. 2u: an exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106(1):77–107, 1998.

[100] Enrico Fagiuoli and Marco Zaffalon. 2U: an exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106:77–107, 1998.

[101] Enrico Fagiuoli and Marco Zaffalon. Tree-augmented naive credal classifiers. In *IPMU 2000: Proceedings of the 8th Information Processing and Management of Uncertainty in Knowledge-Based Systems Conference*, pages 1320–1327, 2000.

[102] Alvaro Eduardo Faria. *Graphical Bayesian models in multivariate expert judgements and conditional external Bayesianity.* PhD thesis, University of Warwick, 1996.

[103] Malcolm Farrow. J. b. paris, the uncertain reasoner's companion: A mathematical perspective. cambridge university press, cambridge, 1994. isbn 0-521-46089-1. *Natural Language Engineering*, 4(4):363–382, 1998.

[104] Kenneth Friedman and Abner Shimony. Jaynes's maximum entropy prescription and probability theory. *Journal of Statistical Physics*, 3(4):381–384, 1971.

[105] Nir Friedman. The bayesian structural em algorithm. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 129–138. Morgan Kaufmann Publishers Inc., 1998.

[106] Akira Fusaoka and Satoshi Hiratsuka. On a linear representation for quantitative belief revision. *Transactions of the Japanese Society for Artificial Intelligence*, 18:75–85, 2003.

[107] David Galles and Judea Pearl. Axioms of causal relevance. *Artificial Intelligence*, 97(1-2):9–43, 1997.

[108] David Galles and Judea Pearl. An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3(1):151–182, 1998.

[109] Peter Gärdenfors. *Knowledge in flux: Modeling the dynamics of epistemic states.* The MIT press, 1988.

[110] Michael R. Garey and David S. Johnson. *Computers and intractability: a guide to the theory of NP-completeness.* W. H. Freeman & Co., 1979.

[111] Ashutosh Garg, TS Jayram, Shivakumar Vaithyanathan, and Huaiyu Zhu. Generalized opinion pooling. In *ISAIM*, 2004.

[112] Dan Geiger and David Heckerman. Knowledge representation and inference in similarity networks and bayesian multinets. *Artificial Intelligence*, 82(1-2):45–74, 1996.

[113] Christian Genest, Kevin J McConway, and Mark J Schervish. Characterization of externally bayesian pooling operators. *The Annals of Statistics*, pages 487–501, 1986.

[114] Christian Genest and Carl G Wagner. Further evidence against independence preservation in expert judgement synthesis. *Aequationes Mathematicae*, 32(1):74–86, 1987.

[115] Christian Genest, Samaradasa Weerahandi, and James V Zidek. Aggregating opinions through logarithmic pooling. *Theory and Decision*, 17(1):61–70, 1984.

[116] Christian Genest and James V Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, pages 114–135, 1986.

[117] Richard D Gill, Mark J Van Der Laan, and James M Robins. Coarsening at random: Characterizations, conjectures, counter-examples. *LECTURE NOTES IN STATISTICS-NEW YORK-SPRINGER VERLAG-*, pages 255–294, 1997.

[118] Moisés Goldszmidt and Judea Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84(1-2):57–112, 1996.

[119] Adam Grove. Two modellings for theory change. *Journal of philosophical logic*, 17(2):157–170, 1988.

[120] Adam J Grove and Joseph Y Halpern. Probability update: Conditioning vs. cross-entropy. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, pages 208–214. Morgan Kaufmann Publishers Inc., 1997.

[121] Peter D Grunwald and Joseph Y Halpern. Updating probabilities. *Journal of Artificial Intelligence Research*, 19:243–278, 2003.

[122] Peng Guan, Rebecca Howell-Jones, Ni Li, Laia Bruni, Silvia de Sanjosé, Silvia Franceschi, and Gary M Clifford. Human papillomavirus types in 115,789 hpv-positive women: a meta-analysis from cervical infection to cancer. *International journal of cancer*, 131(10):2349–2359, 2012.

[123] Mario Günther. Learning conditional information by Jeffrey imaging on Stalnaker conditionals. *Journal of Philosophical Logic*, pages 1–26, 2017.

[124] Giorgio Guzzetta, Luca Faustini, Donatella Panatto, Roberto Gasparini, and Piero Manfredi. The impact of hpv female immunization in italy: model based predictions. *PloS one*, 9(3):e91698, 2014.

[125] Peter J Hammond. Elementary non-archimedean representations of probability for decision theory and games. In *Patrick Suppes: scientific philosopher*, pages 25–61. Springer, 1994.

[126] William L Harper. Rational belief change, popper functions and counterfactuals. *Synthese*, 30(1-2):221–262, 1975.

[127] David Heckerman, Dan Geiger, and David M Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3):197–243, 1995.

[128] Daniel F Heitjan and Donald B Rubin. Ignorability and coarse data. *The annals of statistics*, pages 2244–2253, 1991.

[129] Jon C Helton and Freddie Joe Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23–69, 2003.

[130] Max Henrion. Search-based methods to bound diagnostic probabilities in very large belief nets. In *Uncertainty Proceedings 1991*, pages 142–150. Elsevier, 1991.

[131] Niel Hens, Ziv Shkedy, Marc Aerts, Christel Faes, Pierre Van Damme, and Philippe Beutels. *Modeling infectious disease parameters based on serological and social contact data: a modern statistical perspective*, volume 63. Springer Science & Business Media, 2012.

[132] Andreas Herzig and Omar Rifi. Propositional belief base update and minimal change. *Artificial Intelligence*, 115(1):107–138, 1999.

[133] Leon Horsten and Richard Pettigrew. *The Continuum companion to philosophical logic*. A&C Black, 2011.

[134] Franz Huber. Belief revision i: the agm theory. *Philosophy Compass*, 8(7):604–612, 2013.

[135] Patrick Hurley. *A concise introduction to logic*. Cengage Learning, 2005.

[136] Hidetomo Ichihashi and Hideo Tanaka. Jeffrey-like rules of conditioning for the dempster-shafer theory of evidence. *International Journal of Approximate Reasoning*, 3(2):143–156, 1989.

[137] Ronald L Iman and William-Jay Conover. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics-Simulation and Computation*, 11(3):311–334, 1982.

[138] Richard Jeffrey. Conditioning, kinematics, and exchangeability. In *Causation, chance and credence*, pages 221–255. Springer, 1988.

[139] Richard C Jeffrey. Ethics and the logic of decision. *The Journal of Philosophy*, 62(19):528–539, 1965.

[140] Vomlel Jiří. *Methods of Probabilistic Knowledge Integration*. PhD thesis, Ph. D. Thesis, 1999.

[141] Radim Jiroušek. Solution of the marginal problem and decomposable distributions. *Kybernetika*, 27(5):403–412, 1991.

[142] Radim Jiroušek and Stanislav Přeučil. On the effective implementation of the iterative proportional fitting procedure. *Computational Statistics & Data Analysis*, 19(2):177–189, 1995.

[143] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

[144] Sham M Kakade, Michael Kearns, and Luis E Ortiz. Graphical economics. In *International Conference on Computational Learning Theory*, pages 17–32. Springer, 2004.

[145] Markus Kalisch and Peter Bühlmann. Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *Journal of Machine Learning Research*, 8(Mar):613–636, 2007.

[146] Markus Kalisch, Martin Mächler, Diego Colombo, Marloes H Maathuis, Peter Bühlmann, et al. Causal inference using graphical models with the r package pcalg. *Journal of Statistical Software*, 47(11):1–26, 2012.

[147] Hirofumi Katsuno and Alberto O Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52(3):263–294, 1991.

[148] Kevin T Kelly. Iterated belief revision, reliability, and inductive amnesia. *Erkenntnis*, 50(1):7–53, 1999.

[149] W Kermack and A McKendrick. A contribution to the mathematical theory of epidemics. In *Proc. Roy. Soc. Lond*, pages 700–721, 1927.

[150] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques.* MIT press, 2009.

[151] Andrei Nikolaevich Kolmogorov. Foundations of probability. 1933.

[152] Sébastien Konieczny and Ramón Pino-Pérez. On the logic of merging. 1998.

[153] Rudolf Kruse, Christian Borgelt, Christian Braune, Sanaz Mostaghim, and Matthias Steinbrecher. *Computational intelligence: a methodological introduction.* Springer, 2016.

[154] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[155] Steffen L Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

[156] Steffen L Lauritzen and David J Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 157–224, 1988.

[157] Isaac Levi. The enterprise of knowledge, 1980.

[158] Isaac Levi. *The fixation of belief and its undoing: changing beliefs through inquiry.* Cambridge University Press, 1991.

[159] David Lewis. Probabilities of conditionals and conditional probabilities. In *Ifs*, pages 129–147. Springer, 1976.

[160] David Lewis. On the plurality of worlds. *London*, 5:221–36, 1986.

[161] Zhaoyu Li and Bruce d'Ambrosio. Efficient inference in bayes networks as a combinatorial optimization problem. *International Journal of Approximate Reasoning*, 11(1):55–81, 1994.

[162] Churn-Jung Liau. Belief, information acquisition, and trust in multi-agent systems: a modal logic formulation. *Artificial Intelligence*, 149(1):31–60, 2003.

[163] Paolo Liberatore. The complexity of iterated belief revision. In *International Conference on Database Theory*, pages 276–290. Springer, 1997.

[164] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2014.

[165] Jianbing Ma, Weiru Liu, Didier Dubois, and Henri Prade. Bridging Jeffrey's rule, AGM revision and Dempster conditioning in the theory of evidence. *International Journal on Artificial Intelligence Tools*, 20(04):691–720, 2011.

[166] Albert Madansky. Externally bayesian groups. Technical report, Rand Corp Santa Monica CA, 1964.

[167] David Makinson. General theory of cumulative inference. In *Non-monotonic reasoning*, pages 1–18. Springer, 1989.

[168] Sabina Marchetti. The soft propagation algorithm: A proposal for responsive belief revision with bayesian networks (presented poster at 10th isipta), 2017.

[169] Sabina Marchetti. Informed pooling: an application to the game of italian football team s.s. lazio. Unpublished Manuscript.

[170] Sabina Marchetti and Alessandro Antonucci. Imaginary kinematics. In *Proceedings of the Uncertainty in Artificial Intelligence Conference 2018*, 2018.

[171] Sabina Marchetti and Alessandro Antonucci. Reliable uncertain evidence modeling in Bayesian networks by credal networks. In *Proceedings of the 31st International FLAIRS Conference*, pages 513–518. AAAI, 2018.

[172] Sabina Marchetti, Luca Faustini, Giorgio Guzzetta, Paolo Bonanni, Sara Boccalini, and Piero Mafredi. Conseguenze epidemiologiche e di sanità pubblica della vaccinazione universale con il vaccino hpv 9-valente : predizioni basate su un modello matematico. Technical report, FIRB, 2018.

[173] Sabina Marchetti, Giorgio Guzzetta, Elmira Flem, Grazina Mirinaviciute, Gianpaolo Scalia Tomba, and Piero Manfredi. Modeling the impact of combined vaccination programs against varicella and herpes zoster in norway. *Vaccine*, 36(8):1112–1125, 2018.

[174] Simeone Marino, Ian B Hogue, Christian J Ray, and Denise E Kirschner. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology*, 254(1):178–196, 2008.

[175] D.D. Mauá, C.P. de Campos, A. Benavoli, and A. Antonucci. Probabilistic inference in credal networks: new complexity results. *Journal of Artificial Intelligence Research*, 50:603–637, 2014.

[176] Denis D Mauá, Cassio P de Campos, Alessio Benavoli, and Alessandro Antonucci. On the complexity of strong and epistemic credal networks. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 391–400. AUAI Press, 2013.

[177] Denis D Mauá, Cassio P De Campos, and Marco Zaffalon. Updating credal networks is approximable in polynomial time. *International Journal of Approximate Reasoning*, 53(8):1183–1199, 2012.

[178] Pedrito Maynard-Reid and Yoav Shoham. Belief fusion: Aggregating pedigreed belief states. *Journal of Logic, Language and Information*, 10(2):183–209, 2001.

[179] Kevin J McConway. Marginalization and linear opinion pools. *Journal of the American Statistical Association*, 76(374):410–414, 1981.

[180] Michael D McKay, Richard J Beckman, and William J Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[181] Michael D McKay, Richard J Beckman, and William J Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[182] Danny A Milner. *Diagnostic Pathology: Infectious Diseases E-Book*. Elsevier Health Sciences, 2015.

[183] Enrique Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658, 2008.

[184] Enrique Miranda and Marco Zaffalon. Coherence graphs. *Artificial Intelligence*, 173(1):104–144, 2009.

[185] James Mitchell et al. The recalibrated and copula opinion pools. *Warwick Business School*, 2013.

[186] Serafín Moral and Andrés Cano. Strong conditional independence for credal sets. *Annals of Mathematics and Artificial Intelligence*, 35(1):295–321, 2002.

[187] Peter A Morris. Combining expert judgments: A bayesian approach. *Management Science*, 23(7):679–693, 1977.

[188] Ali Ben Mrad, Véronique Delcroix, Sylvain Piechowiak, Philip Leicester, and Mohamed Abid. An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence. *Applied Intelligence*, 43(4):802–824, 2015.

[189] Richard E Neapolitan et al. *Learning bayesian networks*, volume 38. Pearson Prentice Hall Upper Saddle River, NJ, 2004.

[190] Mark Newman. *Networks: an introduction*. Oxford university press, 2010.

[191] Radu Stefan Niculescu, Tom M Mitchell, and R Bharat Rao. Bayesian network learning with parameter constraints. *Journal of Machine Learning Research*, 7(Jul):1357–1383, 2006.

[192] Hans A Panofsky, Glenn W Brier, and William H Best. Some application of statistics to meteorology. 1958.

[193] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California, 1988.

[194] Judea Pearl. Morgan kaufmann series in representation and reasoning. probabilistic reasoning in intelligent systems: Networks of plausible inference, 1988.

[195] Judea Pearl and Azaria Paz. *Graphoids: A graph-based logic for reasoning about relevance relations*. University of California (Los Angeles). Computer Science Department, 1985.

[196] Arthur Paul Pedersen and Gregory Wheeler. Demystifying dilation. *Erkenntnis*, 79(6):1305–1342, 2014.

[197] Yun Peng, Shenyong Zhang, and Rong Pan. Bayesian network reasoning with uncertain evidences. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 18(05):539–564, 2010.

[198] David M Pennock and Michael P Wellman. Graphical representations of consensus belief. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 531–540. Morgan Kaufmann Publishers Inc., 1999.

[199] Johan Pensar, Henrik Nyman, Timo Koski, and Jukka Corander. Labeled directed acyclic graphs: a generalization of context-specific independence in directed graphical models. *Data mining and knowledge discovery*, 29(2):503–533, 2015.

[200] Alberto Piatti, Marco Zaffalon, Fabio Trojani, and Marcus Hutter. Limits of learning about a categorical latent variable under prior near-ignorance. *International Journal of Approximate Reasoning*, 50(4):597–611, 2009.

[201] Frédéric Pichon and Thierry Denœux. The unnormalized dempster's rule of combination: a new justification from the least commitment principle and some extensions. *Journal of Automated Reasoning*, 45(1):61–87, 2010.

[202] Gabriella Pigozzi. Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese*, 152(2):285–298, 2006.

[203] David Poole and Adrian E Raftery. Inference for deterministic simulation models: the bayesian melding approach. *Journal of the American Statistical Association*, 95(452):1244–1255, 2000.

[204] ST Rachev, W Sun, and M Stein. Copula concepts in financial markets. *Portfolio Institutionell (Ed.)*, 4:12–15, 2009.

[205] Des Raj. *Sampling theory*. New York, 1968.

[206] Raghav Ramachandran, Abhaya C Nayak, and Mehmet A Orgun. Belief erasure using partial imaging. In *Australasian Joint Conference on Artificial Intelligence*, pages 52–61. Springer, 2010.

[207] Marco Ramoni and Paola Sebastiani. Robust learning with missing data. *Machine Learning*, 45(2):147–170, 2001.

[208] Frank P Ramsey. Truth and probability (1926). *The foundations of mathematics and other logical essays*, pages 156–198, 1931.

[209] Joseph Ramsey, Jiji Zhang, and Peter L Spirtes. Adjacency-faithfulness and conservative causal inference. *arXiv preprint arXiv:1206.6843*, 2012.

[210] Gavin Rens and Thomas Meyer. Imagining probabilistic belief change as imaging (technical report). *arXiv preprint arXiv:1705.01172*, 2017.

[211] Gavin Rens, Thomas Meyer, and Giovanni Casini. On revision of partially specified convex probabilistic belief bases. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI-16)*, 2016.

[212] Alfréd Rényi. On a new axiomatic theory of probability. *Acta Mathematica Academiae Scientiarum Hungarica*, 6(3-4):285–335, 1955.

[213] Irina Rish. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46. IBM New York, 2001.

[214] Hans Rott. *Change, choice and inference: A study of belief revision and nonmonotonic reasoning.* Number 42. Oxford University Press, 2001.

[215] Hans Rott. Information structures in belief revision. *Handbook of the Philosophy of Information*, pages 457–482, 2008.

[216] Jeffrey Sanford Russell, John Hawthorne, and Lara Buchak. Groupthink. *Philosophical studies*, 172(5):1287–1309, 2015.

[217] Stuart Russell and Peter Norvig. A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, 25:27, 1995.

[218] Eugene Santos and Solomon Eyal Shimony. Deterministic approximation of marginal probabilities in bayes nets. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 28(4):377–393, 1998.

[219] Vladimir P Savchuk and Harry F Martz. Bayes reliability estimation using multiple sources of prior information: binomial sampling. *IEEE Transactions on Reliability*, 43(1):138–144, 1994.

[220] Marco Scutari and Jean-Baptiste Denis. *Bayesian networks: with examples in R.* CRC press, 2014.

[221] Teddy Seidenfeld. Entropy and uncertainty. *Philosophy of Science*, 53(4):467–491, 1986.

[222] Kari Sentz, Scott Ferson, et al. *Combination of evidence in Dempster-Shafer theory*, volume 4015. Citeseer, 2002.

[223] Ross D Shachter. Bayes-ball: Rational pastime (for determining irrelevance and requisite information in belief networks and influence diagrams). In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 480–487. Morgan Kaufmann Publishers Inc., 1998.

[224] Glenn Shafer. *A mathematical theory of evidence*, volume 42. Princeton university press, 1976.

[225] Claude E Shannon. A mathematical theory of communication, part i, part ii. *Bell Syst. Tech. J.*, 27:623–656, 1948.

[226] Prakash P Shenoy and Glenn Shafer. Axioms for probability and belief-function propagation. *Classic Works of the Dempster-Shafer Theory of Belief Functions*, 219:499–528, 2008.

[227] Ricardo Shirota Filho, Fabio Gagliardi Cozman, Felipe Werndl Trevizan, Cassio Polpo de Campos, and Leliane Nunes de Barros. Multilinear and integer programming for markov decision processes with imprecise probabilities. *ISIPTA'07*, page 395, 2007.

[228] Tomoji Shogenji. Is coherence truth conducive? *Analysis*, 59(264):338–345, 1999.

[229] John Shore and Rodney Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on information theory*, 26(1):26–37, 1980.

[230] Christophe Simon, Philippe Weber, and Eric Levrat. Bayesian networks and evidence theory to model complex systems reliability. *Journal of Computers (JCP)*, 2(1):33–43, 2007.

[231] Philippe Smets. Constructing the pignistic probability function in a context of uncertainty. In *UAI*, volume 89, pages 29–40, 1989.

[232] Philippe Smets. Probability, possibility, belief: Which and where? In *Quantified Representation of Uncertainty and Imprecision*, pages 1–24. Springer, 1998.

[233] Philippe Smets. Belief functions: the disjunctive rule of combination and the generalized bayesian theorem. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*, pages 633–664. Springer, 2008.

[234] Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search.* MIT press, 2000.

[235] Wolfgang Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In *Causation in decision, belief change, and statistics*, pages 105–134. Springer, 1988.

[236] Herbert A Steinberg. Generalized quota sampling. *Nuclear Science and Engineering*, 15(2):142–145, 1963.

[237] Rush T Stewart and Ignacio Ojea Quintana. Learning and pooling, pooling and learning. *Erkenntnis*, pages 1–21, 2017.

[238] Rush T Stewart and Ignacio Ojea Quintana. Probabilistic opinion pooling with imprecise probabilities. *Journal of Philosophical Logic*, pages 1–29, 2017.

[239] M. Stone. The linear pool. *Annals of Mathematical Statistics*, 32:1339–1342, 1961.

[240] Erik Štrumbelj. On determining probability forecasts from betting odds. *International journal of forecasting*, 30(4):934–943, 2014.

[241] Milan Studenỳ. Semigraphoids and structures of probabilistic conditional independence. *Annals of Mathematics and Artificial Intelligence*, 21(1):71–98, 1997.

[242] Alfred Tarski. Fundamental concepts of the methodology of the deductive sciences. *Logic, semantics, metamathematics*, pages 60–109, 1956.

[243] L Terrence. Fine. theories of probability. *AcademicPress, London&New York*, pages 123–125, 1973.

[244] Marco Valtorta, Young-Gyun Kim, and Jiří Vomlel. Soft evidential update for probabilistic multiagent systems. *International Journal of Approximate Reasoning*, 29(1):71–106, 2002.

[245] Linda C van der Gaag and Silja Renooij. Probabilistic networks as probabilistic forecasters. In *Conference on Artificial Intelligence in Medicine in Europe*, pages 294–298. Springer, 2003.

[246] Barbara Vantaggi. Graphical models for conditional independence structures. In *ISIPTA*, volume 1, pages 332–341, 2001.

[247] Barbara Vantaggi. The l-separation criterion for description of cs-independence models. *International Journal of Approximate Reasoning*, 29(3):291–316, 2002.

[248] Barbara Vantaggi. Graphical representation of asymmetric graphoid structures. In *ISIPTA*, volume 3, pages 560–574, 2003.

[249] TS Verma and J Pearl. Equivalence and synthesis of causal models [technical report r-150]. *Department of Computer Science, University of California, Los Angeles*, 1990.

[250] Jiří Vomlel. Integrating inconsistent data in a probabilistic model. *Journal of Applied Non-Classical Logics*, 14(3):367–386, 2004.

[251] Carl Wagner. Probability revision, the uniformity rule, and the chan–darwiche metric. *Soft Methods for Handling Variability and Imprecision*, pages 74–79, 2008.

[252] Carl Wagner. Jeffrey conditioning and external bayesianity. *Logic Journal of IGPL*, 18(2):336–345, 2009.

[253] Carl Wagner and K Lehrer. Rational consensus in science and society. *D. Reidel*, 1981.

[254] Carl G. Wagner. Probability kinematics and commutativity. *Philosophy of Science*, 69(2):266–278, 2002.

[255] Carl G Wagner. Commuting probability revisions: The uniformity rule. *Erkenntnis*, 59(3):349–364, 2003.

[256] Peter Walley. The elicitation and aggregation of beliefs. *Statistics Research Report*, 23, 1982.

[257] Peter Walley. *Statistical reasoning with imprecise probabilities.* Chapman & Hall, 1991.

[258] Peter Walley and Terrence L Fine. Towards a frequentist theory of upper and lower probability. *The Annals of Statistics*, pages 741–761, 1982.

[259] Peter Walley, Renato Pelessoni, and Paolo Vicig. Direct algorithms for checking consistency and making inferences from conditional probability assessments. *Journal of Statistical Planning and Inference*, 126(1):119–151, 2004.

[260] Larry Wasserman and Teddy Seidenfeld. The dilation phenomenon in robust bayesian inference. *Journal of statistical planning and inference*, 40(2-3):345–356, 1994.

[261] Michael P Wellman and Chao-Lin Liu. State-space abstraction for anytime evaluation of probabilistic networks. In *Uncertainty Proceedings 1994*, pages 567–574. Elsevier, 1994.

[262] Jon Williamson. Aggregating judgements by merging evidence. *Journal of Logic and Computation*, 19(3):461–473, 2008.

[263] Jon Williamson. *In defence of objective Bayesianism*, volume 24. Oxford University Press, 2010.

[264] George Wilmers. The social entropy process: Axiomatising the aggregation of probabilistic beliefs. 2010.

[265] SK Michael Wong and Cory J Butz. Contextual weak independence in bayesian networks. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 670–679. Morgan Kaufmann Publishers Inc., 1999.

[266] Hong Xu and Philippe Smets. Reasoning in evidential networks with conditional belief functions. *International Journal of Approximate Reasoning*, 14(2-3):155–185, 1996.

[267] Boutheina Ben Yaghlane and Khaled Mellouli. Inference in directed evidential networks based on the transferable belief model. *International Journal of Approximate Reasoning*, 48(2):399–418, 2008.

[268] Boutheina Ben Yaghlane, Philippe Smets, and Khaled Mellouli. Directed evidential networks with conditional belief functions. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 291–305. Springer, 2003.

[269] L Zadeh. On the validity of dempster's rule of combination, memo m 79/24. *Univ. of California, Berkeley*, 74, 1979.

[270] Marco Zaffalon. Conservative rules for predictive inference with incomplete data. In *ISIPTA*, volume 5, pages 406–415, 2005.

[271] Marco Zaffalon and Enrique Miranda. Conservative inference rule for uncertain reasoning under incompleteness. *Journal of Artificial Intelligence Research*, 34:757–821, 2009.

[272] Marco Zaffalon and Enrique Miranda. Probability and time. *Artificial Intelligence*, 198:1–51, 2013.

[273] Jiji Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17):1873–1896, 2008.

[274] Nevin L Zhang and David Poole. On the role of context-specific independence in probabilistic inference. In *16th International Joint Conference on Artificial Intelligence, IJCAI 1999, Stockholm, Sweden*, volume 2, page 1288, 1999.

[275] Nevin Lianwen Zhang and David Poole. Exploiting causal independence in bayesian network inference. *Journal of Artificial Intelligence Research*, 5:301–328, 1996.

[276] Chunlai Zhou, Mingyue Wang, and Biao Qin. Belief-kinematics Jeffrey's rules in the theory of evidence. In *Proceedings of UAI 2014*, pages 917–926, 2014.

[277] Zhiqiang Zhuang, James Delgrande, Abhaya Nayak, and Abdul Sattar. A unifying framework for probabilistic belief revision. In *Proceedings of IJCAI-17*, 2017.