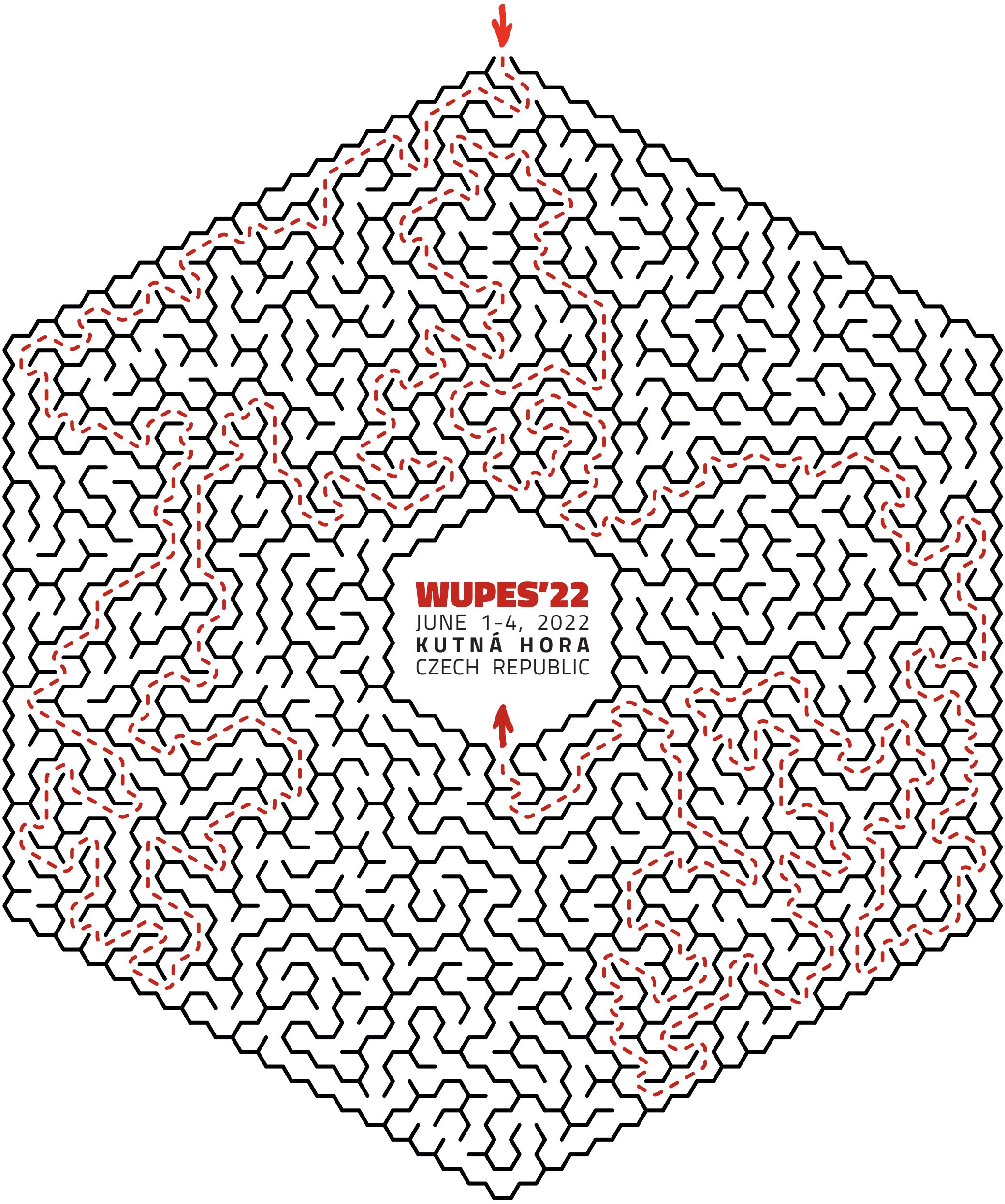
PROCEEDINGS



OF THE 12TH WORKSHOP ON UNCERTAINTY PROCESSING

Proceedings of the 12th Workshop on Uncertainty Processing *(WUPES'22)* Kutná Hora, Czech Republic

Milan Studený, Nihat Ay, Giulianella Coletti, Gernot D. Kleiter, Prakash P. Shenoy
(editors)

June 1-4, 2022

Published by:
MatfyzPress,
Publishing House of the Faculty of Mathematics and Physics Charles University
Sokolovská 83, 186 75 Praha 8, Czech Republic
as the 654. publication.
Printed by Reprostředisko MFF UK.

The text hasn't passed the review or lecturer control of the publishing company MatfyzPress.
The publication has been issued for the purposes of the WUPES'22 conference.
The publishing house Matfyzpress is not responsible for the quality and content of the text.

Printed in Prague — May 2022

Organized by:
Institute of Information Theory and Automation, Czech Academy of Sciences
Faculty of Management, Prague University of Economics and Business

Credits:
Editors: Milan Studený, Nihat Ay, Giulianella Coletti, Gernot D. Kleiter, Prakash P. Shenoy
LaTeXeditor: Václav Kratochvíl
Cover design: Jiří Přibil
using LaTeX's 'confproc' package, version 0.8 by V. Verfaille

ISBN: 978-80-7378-460-7

12th WORKSHOP ON UNCERTAINTY PROCESSING

Wupes '22

KUTNÁ HORA

Organized by:

Institute of Information Theory and Automation,
Czech Academy of Sciences

&

Faculty of Management,
Prague University of Economics and Business

Kutná Hora

June 1-4, 2022

Programme and Conference Committee:
Milan Studený - chair, *Czech Academy of Sciences, Prague*
Nihat Ay, *Hamburg University of Technology, Germany*
Giulianella Coletti, *Università di Perugia, Italy*
Gernot D. Kleiter, *Universität Salzburg, Austria*
Prakash P. Shenoy, *University of Kansas, USA*

Organizing Committee:
Jirka Vomlel - chair, *Czech Academy of Sciences, Prague*
Radim Jiroušek, *Czech Academy of Sciences, Prague*
Václav Kratochvíl, *Czech Academy of Sciences, Prague*
Milan Studený, *Czech Academy of Sciences, Prague*

Foreword

A series of Workshops on Uncertainty Processing, known under its abbreviation WUPES, has been held in Czechia (the first two in Czechoslovakia) every third year since 1988. So, in 2022, it is the 12th edition of the WUPES workshop, which has been postponed one year due to the Covid-19 pandemics. We are very pleased we can meet face to face again and enjoy shared moments during the workshop presentations, discussions, and during social activities of the workshop.

It is a part of the tradition that work in progress stimulating discussions and lectures offering new approaches to research problems are presented at the workshop. Nineteen papers were accepted for presentation at the workshop and are included in these proceedings. The papers cover diverse topics like possibility theory, belief functions, decision making under uncertainty, theoretical foundations and applications of probabilistic graphical models, epidemic models, fuzzy sets, and data mining.

Following its tradition, the workshop takes place in Czechia where it has not been before. This time it is Kutná Hora, a town that played an important role in Czech history due to its silver mines and the royal mint. Once the second-largest city of the Kingdom of Bohemia after the Royal seat of Prague, it is now a quiet town with about 20,000 inhabitants. However, the rich history of the town has left its marks - starting with the Royal Mint established as early as in the year 1300 by the Bohemian King Wenceslas II issuing the mining legislation "Ius regale montanorum and introducing a new silver coin, the so-called Prague groschen. Two historical monuments are of worldwide importance – St. Barbora's Cathedral and the Cathedral of the Assumption of the Virgin Mary. The uniqueness of Kutná Hora was recognized in 1995 when the city was inscribed on the UNESCO World Cultural and Heritage List.

This workshop is co-organized by the two institutions – the Institute of Information Theory and Automation (ÚTIA) of the Czech Academy of Sciences, and the Faculty of Management, Prague University of Economics and Business. We are grateful for their support and the support provided by the Czech National Science Foundation under grants no 19-06569S, 19-04579S, and 20-18407S. I also want to thank all the members of the program and organizing committees.

I wish the participants a pleasant conference that will provide stimuli for their further research. I wish all of us a pleasant time in Kutná Hora.

In Prague, May, 9, 2022

Jirka Vomlel

List of Content

Characterizing Uncertainty In Decision-Making Models For Maintenance In Industry 4.0

Umair Ahmed[1], Silvia Carpitella[2], and Antonella Certa[1]

[1]Department of Engineering, University of Palermo, Viale delle Scienze, 90128, Palermo, Italy
{umair.sfc84,antonella.certa}@unipa.it
[2]Department of Decision-Making Theory, Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou Věží 4, 18208, Prague, Czech Republic;
carpitella@utia.cas.cz

Abstract

Decision-making involves our daily life at any level, something that entails uncertainty and potential occurrence of risks of varied nature. When dealing with industrial engineering systems, effective decisions are fundamental in terms of maintenance planning and implementation. Specifically, several forms of uncertainty may affect decision-making procedures, for which adopting suitable techniques seems to be a good strategy to attain the main maintenance goals by taking into account system criticality along with decision-maker(s) opinions. A wide variety of factors contributes to uncertainty, being some of them greatly important while other ones less significant. However, all of these factors in synergy can impact the functioning of systems in a positive, neutral, or negative way. In this case, the question is whether obtaining a complete picture of such uncertainty can improve decision-making capabilities and mitigate both through-life costs and unforeseen problems. The fundamental issues include dealing with ambiguity in the maintenance decision-making process by employing numerous evaluation criteria and dealing with real-world scenarios in the maintenance environment. In this study, the Multi-Criteria Decision-Making (MCDM) approach is analysed, with particular reference to the Fuzzy Technique for Order of Preference by Similarity to Ideal Solution (FTOPSIS), technique capable to effectively rank alternatives while dealing with uncertainty for maintenance decision-making. A final case study is developed to demonstrate the applicability of the method to the field of maintenance in industry 4.0. The proposed study may be useful in supporting intelligent and efficient decisions resulting in favorable maintenance outcomes.

1 Introduction

Enabled by such industry 4.0 technologies as machine learning, big data, and augmented reality, the current digital era is generally characterised by an abundance of information available to aid in decision-making. Assets can be easily and real-time connected via networks of suitable sensors, commonly referred to as the Internet of Things (IoT). The primary problem has shifted from obtaining data to making educated decisions on the basis of the acquired information. The whole maintenance management relies on such information, as well as on how to utilise data and predictive analytic to improve our judgments. As a consequence, new possibilities for data-driven techniques including predictive analytic, artificial intelligence, and machine learning have been developed, with the potential for large efficiency advantages. Everyday life is associated with constant decision-making and each of these decisions involves potential of uncertainty and risk (Van Staden, 2021), something that can directly influence maintenance strategies.

Numerous variables contribute to uncertainty, some of them are extremely significant while other ones may be inconsequential, affecting performance of the system in a favourable, neutral, or negative way (Grenyer et al., 2019). On the whole, two different forms of uncertainty can be distinguished: quantitative, based on recorded statistical data, and qualitative, based on unobserved statistical data consisting of heuristic estimates obtained from expert opinions, supplier specifications, and equipment accuracy. On the one hand, the first category is well-documented and can be simply represented as the standard deviation of a particular data set. On the other hand, the second category is often difficult to be characterised. Additionally, uncertainty can be classified as epistemic and aleatory. The first one stems from model or data accuracy, which is impacted by the available amount of knowledge, and may therefore be alleviated or improved. The other type denotes statistical variables that change continually and so cannot be minimized (Grenyer et al., 2019). Among the several causes of uncertainty, the primary source is the lack of knowledge about engineering phenomena. Indeed, decision-making processes are affected by several types of uncertainty, depending on its own root causes.

Uncertainty manifests itself at several levels in diagnostic problems, particularly when it comes to information and/or system defects. Two primary aspects of uncertainty refers to the available information used to support decision-making problems: fuzziness and stochasticity. The ideal decision-making procedures under situations of uncertainty in order to achieve the maintenance objective vary according to the system's nature and the decision maker's priorities (Borissova et al., 2011). Currently, industrial maintenance decision-making is primarily based on two major categories of data: captured data and subjective expert views. The collected data contains objective facts that are subjected to a degree of uncertainty statistically quantifiable as the standard deviation of the dataset under analysis. Subjective expert opinions assign qualitative uncertainty to individuals based on their characteristics that qualify them as experts and the foundation for their perspective in order to prove its legitimacy. The precision of the equipment employed together with the competence of the maintainer are seldom recognised as contributors to total uncertainty in methods of data collection. However, their roles are fundamental to comprehensively characterise and manage uncertainty, as exemplified in Figure 1.

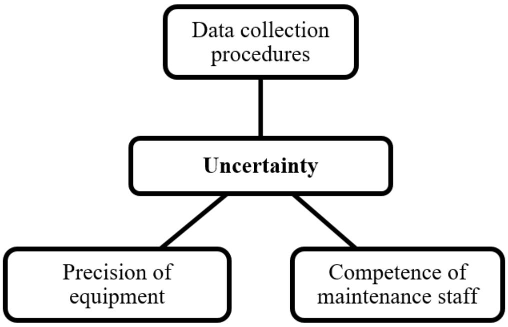


Figure 1: Contributors of uncertainty in maintenance (Grenyer et al., 2019)

A mix of objective data and subjective opinions should be considered to elicit reliable judgments leading to effective maintenance results. Certain instances need indeed further skills while other ones necessitate additional data. The issue is whether taking a comprehensive picture of such uncertainties may help to enhance decision-making capability and mitigate both through-life costs and unanticipated problems (Grenyer et al., 2019).

Reviewing and adapting maintenance policies to the many possibilities available in systems or plants is critical for maintenance managers. Especially when multiple conflicting criteria and methods are taken into account, it is difficult to undertake proper maintenance strategies. The primary problems include dealing with uncertainty in the evaluation of maintenance policies using multiple assessment criteria and dealing with real-world situations in maintenance (Mojtahedi et al., 2020). In this study, we are going to assume a Multi-Criteria Decision-Making (MCDM) perspective and, in particular, an approach based on the Fuzzy Technique for Order of Preference by Similarity to Ideal Solution (FTOPSIS) is going to be applied to rank alternatives relevant to industry 4.0 in order to characterise uncertainty in maintenance decision-making. The proposed study may be useful in supporting companies to make effective decisions optimising business results on the whole.

2 Literature review

MCDM methods are extensively implemented in many domains, e.g. engineering, supply chain management, economics, social sciences, medical sciences, among others. Despite its variety, the MCDM paradigm shares several aims and criteria that are sometimes in conflict with each other. Over the last decades, MCDM methods have grown in importance in such fields as operations research (Nădăban et al., 2016), and their adoption is commonly considered to be a robust scientific strategy to make intelligent and acceptable decisions in complex maintenance contexts (Abdulgader et al., 2018) such as those involved in industry 4.0. Various MCDM methodologies have been largely used by several professionals in different areas of study (Palczewski and Sałabun, 2019).

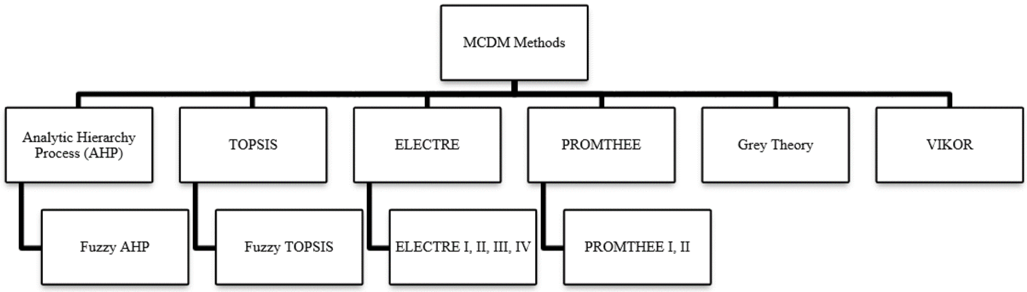


Figure 2: MCDM techniques and types (Aruldoss et al., 2013)

Some of these techniques are summarised by Aruldoss et al. (2013), as recalled in Figure 2, and can be applied in their traditional version or even in their fuzzy developments. In the first case, decision-making elements (i.e. criteria, sub-criteria, alternatives) are evaluated, ranked and/or weighted on the basis of assessments given in the form of crisp numbers. Alternatively, in the second case, linguistic variables to be translated into fuzzy numbers are used in order to better manage the ambiguity as well as the lack of precision and clarity affecting input evaluations (Wang and Lee, 2009).

Among the MCDM methods available in literature, we are going to discuss about the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) along with its fuzzy extension (FTOPSIS). This choice is justified by the fact that these techniques allow extreme flexibility in ranking elements, something that appears to be particularly useful in modern maintenance contexts, greatly impacted by digital transformations.

2.1 Traditional TOPSIS: advantages and limitations

In the vast majority of real-life scenarios, given the ambiguity of human preference behaviour, decision-makers are often unable to produce effectively representative numerical evaluations for discriminating among the main elements of a complex problem. Numerous MCDM approaches have been developed and applied over the years and, among them, TOPSIS is one of the most common methods used in literature to deal with complex decision-making problems (Salih et al., 2019; Palczewski and Sałabun, 2019; Hung and Chen, 2009; Kutlu and Ekmekçioğlu, 2012; Kore et al., 2017), with the ultimate goal of producing a structured ranking of alternatives (Kutlu and Ekmekçioğlu, 2012; Gupta, 2018) on the basis of evaluation criteria, suitably weighted.

TOPSIS was established on the notion that the selected alternative(s) should have the shortest distance to an ideal point, called Positive Ideal Solution (PIS) and, simultaneously, the longest distance to another ideal point, called Negative Ideal Solution (NIS) (Wang and Lee, 2009; Hung and Chen, 2009; Kutlu and Ekmekçioğlu, 2012; Kore et al., 2017; Wang and Elhag, 2006). The output is then based on the calculation, for each alternative, of the positive and negative distances (Solangi et al., 2021). To such an aim, an accommodative aggregation technique may preliminary evaluate a set of alternatives by assigning weights to each criterion (Palczewski and Sałabun, 2019). However, using actual crisp values to score the alternatives under analysis may lead to restrictions in addressing uncertainty (Salih et al., 2019). In any case, TOPSIS includes an easily comprehensible and flexible calculation technique having the capability to take into consideration several criteria with varied units at the same time (Kutlu and Ekmekçioğlu, 2012). Given to its great flexibility of application, TOPSIS is a prominent MCDM method employed by many scholars in a huge variety of sectors (Solangi et al., 2021; Behzadian et al., 2012). Moreover, it has been widely integrated with several other MCDM strategies as an efficient way for prioritizing maintenance decision-making (see Singh et al. (2016); Ighravwe and Oke (2021), among others). TOPSIS is indeed considered to be faster, apart from much more adaptable, comprehensible, and straightforward than many other MCDM methodologies (Haddad et al., 2021).

TOPSIS' strengths comprise transparency, intuitively grasped concepts, improved working efficiency, and capability to evaluate the overall efficiency of each alternative in a simple mathematical format, something that has resulted in the broad acceptance and understanding of this approach from a varied range of industries (Hung and Chen, 2009). The main benefit of employing TOPSIS is that it requires just few data sets from professionals, such as criteria values and linguistic evaluations of alternatives (Gupta, 2018). It accepts contributions in the form of any set of criteria and characteristics. Because of the notion of detachment from flawless patterns, it has actually instinctual physical significance. It is indeed extremely effective in dealing with circumstances in which maintenance managers, due to their specialized knowledge, believe that technical difficulties may be scaled from the most significant to the least critical considerations. The discussed peculiarities of TOPSIS make it a viable choice for dealing with prioritization issues (Ighravwe and Oke, 2021), also considering the possibility to take simultaneously into account optimal and critical solutions by means of an easy mathematical programming procedure (Rani et al., 2020).

Despite its widespread use, TOPSIS has several limitations in its traditional form, since it actually fails to offer precise information when problems are particularly ambiguous and unexplained (Solangi et al., 2021). Additionally, the use of crisp values for evaluating alternatives is generally inefficient in capturing the subjective character of human cognition. This may lead the technique to fail in effectively reflecting decision makers' priorities in real-world scenarios (Haddad et al., 2021). In multi-criteria contexts, variables are usually in discordant proportions, something that generates complex assessment challenges. Furthermore, TOPSIS' weaknesses may originate the following flaws: (1) its simplistic application may produce incorrect findings; (2) its traditional deterministic version may not exhaustively help in considering uncertainty (Abdulgader et al., 2018). As a result, standard TOPSIS can only partially accommodate ambiguous or vague input through expert opinions. To address all of the mentioned shortcomings, various works of research have integrated fuzzy logic ideas within MCDM approaches. In such a direction, the FTOPSIS technique, originally deveoped by Chen (2000), is proposed as a combination of fuzzy set theory and traditional TOPSIS, under which fuzzy values are employed to provide preference ratings by experts (Palczewski and Sałabun, 2019; Salih et al., 2019; Gupta, 2018; Haddad et al., 2021).

2.2 FTOPSIS: effective treatment of uncertainty

In complex decision-making situations related to maintenance management in industry 4.0, analysing the many variables and factors can be a complex task. As we have already explained, extending traditional models to fuzzy logic can significantly help to mitigate this problem, as it has been successfully demonstrated in many industrial applications (Palczewski and Sałabun, 2019). In 1965, Zadeh developed the concept of fuzzy sets for stimulating spontaneous reasoning by taking into account human ambiguity and subjectivity. As the primary goal of fuzzy logic is to grasp the inaccuracy of human thinking and describe it mathematically (Hung and Chen, 2009; Solangi et al., 2021), linguistic variables can be represented by means of fuzzy numbers with an associated degree of membership $\mu(x)$, varying between 0 and 1. Several researchers have been focusing on the possibility to deal with complex uncertain decision-making problems utilizing fuzzy sets theory. Furthermore, in 1993, Gau and Buehrer introduced the concept of ambiguous sets, stressing as a single value cannot testify to its reality (Hung and Chen, 2009).

FTOPSIS is particularly effective in handling ambiguity and uncertainty affecting input data as it results from human perception and evaluation. Given the ambiguity and lack of knowledge in MCDM, linguistic terms used in FTOPSIS can represent inaccurate data so as to better deal with unclear information (Palczewski and Sałabun, 2019; Salih et al., 2019). Indeed, the use of fuzzy numbers for criteria evaluation streamlines the whole assessment process by also making decision-makers more comfortable in expressing their personal opinions when it comes to qualitative criteria. As a result, FTOPSIS represents a simple, practical forecasting and compensatory method to accept or reject potential options based on hard cut-offs (Kore et al., 2017; Wang and Elhag, 2006). However, it is vital to underline that most of the information gathered and used in FTOPSIS derives from human evaluations, something that makes the estimation of values of importance and also strictly dependent on the quantity of data, that hence need to be "dependable, reliable, constant, certain, authentic, real, and respectable". Despite these drawbacks, FTOPSIS can be regarded as an appropriate method to analyse the values and rank relevant decision-making elements on the basis of linguistic variables and related fuzzy numbers (Solangi et al., 2021).

Numerous studies on FTOPSIS and its integrations are identified in literature. Hwang et al. (2022) assessed maintenance criteria for railroad electrical facility systems based on subjective judgment information of decision-makers by using Design Structure Matrix (DSM) and FTOPSIS approaches. Alshraideh et al. (2021) used a FTOPSIS model to identify the most suitable maintenance contractor under unpredictable conditions, by evaluating proposals' quality. Momeni et al. (2011) proposed the FTOPSIS as a tool for selecting maintenance plans by translating uncertain and imprecise judgment from the decision makers into fuzzy figures. Selim et al. (2016) created a maintenance planning framework integrating the FTOPSIS and the Failure Mode and Effect Analysis (FMEA) approaches for determining the repair priorities of the machines in order to decrease and stabilize maintenance expenditures. Chen et al. (2020) applied the FTOPSIS technique to rate and prioritize paths to e-waste implementation management solutions in Ghana while accounting for the subjectivity of decision-maker preferences.

FTOPSIS have been developed to deal with any type of problem, examples are: assessing and prioritizing strategies for long-term deployment of renewable energy technologies in Pakistan (Solangi et al., 2021); evaluating many alternatives against subjective criteria and weighting all of the factors for robot selection (Chu and Lin, 2003); evaluating suppliers under Health Safety and Environment (HSE) criteria in the oil and gas sector to prioritize operations and maintenance contracts (Haddad et al., 2021); and so on. As reported by Kutlu and Ekmekçioğlu (2012), FTOPSIS has been used also for dealing with the following problems: selection of plant location, supplier selection, industrial robotic system selection, municipal solid waste disposal method and site selection, selection of the best energy technology alternative, and modeling consumer product adoption processes.

3 Methodological procedure

As mentioned in some previous works (Brentan et al., 2021; Carpitella et al., 2018), the most common types of fuzzy numbers are Triangular Fuzzy Numbers (TFNs) $\tilde{n}$, herein considered, which can be expressed as follows (Klir and Yuan, 1996):

$$\tilde{n} = (a, b, c); \tag{1}$$

where $a \leq b \leq c$. Common algebraic operations involving one or more fuzzy numbers can be easily performed. For instance, one can write the following equations:

$$\tilde{n}_1 \oplus \tilde{n}_2 = (a_1 + a_2, b_1 + b_2, c_1 + c_2); \tag{2}$$

$$\tilde{n}_1 \odot \tilde{n}_2 = (a_1 \times a_2, b_1 \times b_2, c_1 \times c_2); \tag{3}$$

$$\tilde{n}_1^{-1} = (\frac{1}{c_1}, \frac{1}{b_1}, \frac{1}{a_1}). \tag{4}$$

On the basis of these preliminaries, we now describe the steps needed to implement the FTOPSIS approach (Youssef, 2020; Akram and Arshad, 2019; Ilyas et al., 2021).

- Defining the fuzzy decision matrix $\tilde{X}$ collecting the whole set of input data:

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{1n} \\ \vdots & \ddots & \vdots \\ \tilde{x}_{m1} & \cdots & \tilde{x}_{mn} \end{bmatrix}. \tag{5}$$

  The generic TFN $\tilde{x}_{ij}$ of matrix $\tilde{X}$ corresponds to the rating of alternative i under criterion j:

$$\tilde{x}_{ij} = (a_{ij}, b_{ij}, c_{ij}). \tag{6}$$

- Weighting and normalising matrix $\tilde{X}$ with relation to different criteria, obtaining matrix $\tilde{U}$, whose components are calculated as:

$$\tilde{u}_{ij} = \Big(\frac{a_{ij}}{c_j^*}, \frac{b_{ij}}{c_j^*}, \frac{c_{ij}}{c_j^*}\Big) \times w_{ij}, j \in I'; \tag{7}$$

$$\tilde{u}_{ij} = \Big(\frac{a_j^-}{c_{ij}}, \frac{a_j^-}{b_{ij}}, \frac{a_j^-}{a_{ij}}\Big) \times w_{ij}, j \in I''; \tag{8}$$

$I^{'}$ being the subset of criteria to be maximized, $I^{''}$ the subset of criteria to be minimized, w_j the weight of criterion j and c_j^* and a_j^- calculated as:

$$c_j^* = \max_i c_{ij} \quad if \quad j \in I'; \tag{9}$$

$$a_j^- = \min_i a_{ij} \quad if \quad j \in I''. \tag{10}$$

- Computing distances between each alternative and the fuzzy ideal solutions A^* and A^-:

$$A^* = (\tilde{u}_1^*, \tilde{u}_2^*, \ldots, \tilde{u}_n^*); \tag{11}$$

$$A^- = (\tilde{u}_1^-, \tilde{u}_2^-, \ldots, \tilde{u}_n^-). \tag{12}$$

where $\tilde{u}_j^* = (1, 1, 1)$ and $\tilde{u}_j^- = (0, 0, 0)$, $j = 1 \ldots n$. Distances between each alternative and these ideal points can be computed through the vertex method (Chen, 2000), for which the distance $d(\tilde{m}, \tilde{n})$ between two TFNs $\tilde{m} = (m_1, m_2, m_3)$ and $\tilde{n} = (n_1, n_2, n_3)$ corresponds to the crisp value:

$$d(\tilde{m}, \tilde{n}) = \sqrt{\frac{1}{3}\Big[\big(m_1 - n_1\big)^2 + \big(m_2 - n_2\big)^2 + \big(m_3 - n_3\big)^2\Big]}. \tag{13}$$

Then, aggregating with respect to the whole set of criteria, the distances of each alternative i from A^* and A^- are, respectively:

$$d_i^* = \sum_{j=1}^{n} d(\tilde{u}_{ij}, \tilde{u}_j^*) \quad i = 1, \ldots, n; \tag{14}$$

$$d_i^- = \sum_{j=1}^{n} d(\tilde{u}_{ij}, \tilde{u}_j^-) \quad i = 1, \ldots, n. \tag{15}$$

- Calculating the closeness coefficient CC_i:

$$CC_i = \frac{d_i^-}{d_i^- + d_i^*} \tag{16}$$

To get the final ranking of alternatives it is necessary to order the values of the closeness coefficient related to each alternatives in a decreasing way.

4 Application and discussion

The present case study applies the FTOPSIS technique to rank a set of 13 alternatives, that are the maintenance factors relevant for industry 4.0 identified and formalised in (Ahmed et al., 2022). The considered factors aim to contemplate the role of maintenance digitalization and their final ranking highlights those aspects to be taken primarily into account when planning industrial strategies while considering uncertainty of evaluations. Alternatives have been evaluated under three main criteria, that are safety & security (C_1), process quality (C_2) and cost efficiency (C_3), all of them to be maximised and, in the present application, equally weighted. Linguistic evaluations reported in Table 1 refer to a real company operating in the waste management sector, having been attributed in cooperation with the human resources in charge, respectively, of the maintenance function and of the safety and security system. The used linguistic variables and related TFNs are: VL (1,1,3), very low impact; L (1,3,5), low impact; M (3,5,7), medium impact; H (5,7,9), high impact; VH (7,9,9), very high impact. Table 1 summarises the results of the FTOPSIS application along with the final ranking of maintenance factors.

ID	**Maintenance Factors**	C_1	C_2	C_3	d_i^*	d_i^-	CC_i	**Rank. pos.**
MF_1	Management commitment and support	M	M	M	0.5844	2.4512	0.1925	9^{th}
MF_2	Smart technology development	M	H	M	0.6558	2.3773	0.2162	7^{th}
MF_3	Organizational growth	M	M	H	0.6558	2.3773	0.2162	7^{th}
MF_4	Development of skilled workforce	VH	VH	H	0.8874	2.1277	0.2943	1^{st}
MF_5	Resources required for digitalization	VH	VH	M	0.8160	2.2015	0.2704	3^{rd}
MF_6	Maintenance strategy development	H	H	VH	0.8431	2.1787	0.2790	2^{nd}
MF_7	Corporate culture	M	M	L	0.5161	2.5251	0.1697	10^{th}
MF_8	Change in working practices	M	M	M	0.5161	2.5251	0.1697	10^{th}
MF_9	Effective maintenance system	H	H	H	0.7987	2.2296	0.2637	4^{th}
MF_{10}	Regulatory compliance	M	H	L	0.5875	2.4512	0.1933	8^{th}
MF_{11}	Safety and health awareness	VH	H	M	0.7716	2.2525	0.2552	5^{th}
MF_{12}	Data privacy and security	L	M	M	0.5161	2.5251	0.1697	10^{th}
MF_{13}	Sustainable performance improvement	M	H	H	0.7273	2.3035	0.2400	6^{th}

Table 1: Evaluation of maintenance factors relevant to industry 4.0

By observing Table 1, factor MF_4, that is “development of skilled workforce”, has prominent importance in maximising all the considered criteria, according to the perceptions of the involved experts. It can be noticed that also MF_6 (“maintenance strategy development”) and MF_5 (“resource required for digitalization”) are regarded as priority aspects. On the contrary, factors MF_7, MF_8 and MF_{12} that are, respectively, “corporate culture”, “change in working practice” and “data privacy and security” occupy the last position of the ranking, having associated lower impact with respect to the other maintenance factors. Some of the factors occupy the same position in the ranking, e.g. factors MF_2 and MF_3, and the reason of it is that criteria have associated the same weight. If weights varied, so would do the ranking position. For example, again in the case of MF_2 and MF_3, if higher weight was attributed to the quality criterion and lower weight to the cost efficiency, MF_2 would eventually occupy a higher position in the final ranking with respect to MF_3, the last one having associated lower evaluation under C_2.

5 Conclusions

This work discusses how to deal with uncertainty affecting decision-making processes with a special focus on industry 4.0 maintenance management. After a comprehensive review on MCDM approaches implemented in the field under study, we underline the valuable support provided by the integration of such tools as the fuzzy set theory for managing complex real situations in which uncertain human opinions are elicited. Specifically, we analyse the TOPSIS and FTOPSIS techniques on the basis of their high methodological flexibility, by formalising weaknesses and advantages of both approaches. FTOPSIS reveals to be particularly useful for treating uncertainty, as it can be demonstrated by several applications. After describing methodological details, We implement a real case study aimed at providing helpful practical insights for maintenance managers in the complex era of digital transformation. Future lines of research may refer to the integration of other MCDM methods supporting in a more precise calculation of criteria weights along with useful mathematical tools as, for instance, the probability theory.

References

F. S. Abdulgader, R. Eid, and B. Daneshvar Rouyendegh. Development of decision support model for selecting a maintenance plan using a fuzzy MCDM approach: A theoretical framework. *Applied Computational Intelligence and Soft Computing*, 2018, 2018.

U. Ahmed, S. Carpitella, and A. Certa. Digital transformation in maintenance management. In *27^{th} Summer School Francesco Turco, under review.* AIDI-Italian Association of Industrial Operations Professors, 2022.

M. Akram and M. Arshad. A novel trapezoidal bipolar fuzzy TOPSIS method for group decision-making. *Group Decision and Negotiation*, 28(3):565–584, 2019.

M. Alshraideh, S. Ababneh, E. E. Gunay, and O. Al-Araidah. A fuzzy-TOPSIS model for maintenance outsourcing considering the quality of submitted tender documents. *Eksploatacja i Niezawodność*, 23(3), 2021.

M. Aruldoss, T. M. Lakshmi, and V. P. Venkatesan. A survey on multi criteria decision making methods and its applications. *American Journal of Information Systems*, 1(1): 31–43, 2013.

M. Behzadian, S. K. Otaghsara, M. Yazdani, and J. Ignatius. A state-of-the-art survey of TOPSIS applications. *Expert Systems with applications*, 39(17):13051–13069, 2012.

D. Borissova, I. Mustakerov, and V. Grigorova. Engineering systems maintenance by optimal decision making strategies under uncertainty conditions. *Problems of Eng. Cybernetics & Robotics*, 63:14–21, 2011.

B. M. Brentan, S. Carpitella, J. Izquierdo, E. Luvizotto Jr, and G. Meirelles. District metered area design through multicriteria and multiobjective optimization. *Mathematical Methods in the Applied Sciences*, 2021.

S. Carpitella, S. J. Ocaña-Levario, J. Benítez, A. Certa, and J. Izquierdo. A hybrid multicriteria approach to GPR image mining applied to water supply system maintenance. *Journal of Applied Geophysics*, 159:754–764, 2018.

C.-T. Chen. Extensions of the topsis for group decision-making under fuzzy environment. *Fuzzy sets and systems*, 114(1):1–9, 2000.

D. Chen, D. Faibil, and M. Agyemang. Evaluating critical barriers and pathways to implementation of e-waste formalization management systems in Ghana: a hybrid BWM and fuzzy TOPSIS approach. *Environmental Science and Pollution Research*, 27(35): 44561–44584, 2020.

T.-C. Chu and Y.-C. Lin. A fuzzy TOPSIS method for robot selection. *The International Journal of Advanced Manufacturing Technology*, 21(4):284–290, 2003.

A. Grenyer, F. Dinmohammadi, J. A. Erkoyuncu, Y. Zhao, and R. Roy. Current practice and challenges towards handling uncertainty for effective outcomes in maintenance. *Procedia CIRP*, 86:282–287, 2019.

H. Gupta. Assessing organizations performance on the basis of GHRM practices using BWM and Fuzzy TOPSIS. *Journal of environmental management*, 226:201–216, 2018.

A. N. Haddad, B. B. da Costa, L. S. de Andrade, A. Hammad, and C. A. Soares. Application of Fuzzy-TOPSIS Method in Supporting Supplier Selection with Focus on HSE Criteria: A Case Study in the Oil and Gas Industry. *Infrastructures*, 6(8):105, 2021.

C.-C. Hung and L.-H. Chen. A fuzzy TOPSIS decision making model with entropy weight under intuitionistic fuzzy environment. In *Proceedings of the international multiconference of engineers and computer scientists*, volume 1, pages 13–16. IMECS Hong Kong, 2009.

S. Hwang, J. Kim, H. Kim, H. Kim, and Y. Kim. Suggestion of Maintenance Criteria for Electric Railroad Facilities Based on Fuzzy TOPSIS. *CMC-Computers Materials & Continua*, 70(3):5453–5466, 2022.

D. E. Ighravwe and S. A. Oke. Applying Fuzzy Multi-criteria Decision Making Framework in Evaluating Maintenance Systems with Emphasis on Human Tasks and Errors. *Mahasarakham International Journal of Engineering Technology*, 7(1):67–77, 2021.

M. Ilyas, S. Carpitella, and E. Zoubir. Designing supplier selection strategies under COVID-19 constraints for industrial environments. *Procedia CIRP*, 100:589–594, 2021.

G. J. Klir and B. Yuan. Fuzzy sets and fuzzy logic: theory and applications. *Possibility Theory versus Probabability Theory*, 32(2):207–208, 1996.

N. B. Kore, K. Ravi, and S. Patil. A simplified description of fuzzy TOPSIS method for multi criteria decision making. *International Research Journal of Engineering and Technology*, 4(5):2047–2050, 2017.

A. C. Kutlu and M. Ekmekçioğlu. Fuzzy failure modes and effects analysis by using fuzzy TOPSIS-based fuzzy AHP. *Expert Systems with Applications*, 39(1):61–67, 2012.

M. Mojtahedi, S. M. Mousavi, H. Gitinavard, and N. Foroozesh. Maintenance policy selection considering resilience engineering by a new interval-valued fuzzy decision model under uncertain conditions. *Int. Journal of Science and Technology*, 28(1), 2020.

M. Momeni, M. R. Fathi, and M. K. Zarchi. Fuzzy TOPSIS-based approach to maintenance strategy selection: A case study. 2011.

S. Nădăban, S. Dzitac, and I. Dzitac. Fuzzy TOPSIS: A general view. *Procedia computer science*, 91:823–831, 2016.

K. Palczewski and W. Sałabun. The fuzzy TOPSIS applications in the last decade. *Procedia Computer Science*, 159:2294–2303, 2019.

P. Rani, A. R. Mishra, A. Mardani, F. Cavallaro, M. Alrasheedi, and A. Alrashidi. A novel approach to extended fuzzy TOPSIS based on new divergence measures for renewable energy sources selection. *Journal of Cleaner Production*, 257:120352, 2020.

M. M. Salih, B. Zaidan, A. Zaidan, and M. A. Ahmed. Survey on fuzzy TOPSIS state-of-the-art between 2007 and 2017. *Computers & Operations Research*, 104:207–227, 2019.

H. Selim, M. G. Yunusoglu, and Ş. Yılmaz Balaman. A dynamic maintenance planning framework based on fuzzy TOPSIS and FMEA: application in an international food company. *Quality and Reliability Engineering International*, 32(3):795–804, 2016.

R. K. Singh, A. Gupta, A. Kumar, and T. A. Khan. Ranking of barriers for effective maintenance by using TOPSIS approach. *Quality in Maintenance Engineering*, 2016.

Y. A. Solangi, C. Longsheng, and S. A. A. Shah. Assessing and overcoming the renewable energy barriers for sustainable development in Pakistan: An integrated AHP and fuzzy TOPSIS approach. *Renewable Energy*, 173:209–222, 2021.

H. Van Staden. Decision making in the presence of uncertainty: Industry 4.0 enabled preventive maintenance. 2021.

T.-C. Wang and H.-D. Lee. Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. *Expert systems with applications*, 36(5):8980–8985, 2009.

Y.-M. Wang and T. M. Elhag. Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment. *Expert systems with applications*, 31(2):309–319, 2006.

A. E. Youssef. An integrated MCDM approach for cloud service selection based on TOPSIS and BWM. *IEEE Access*, 8:71851–71865, 2020.

A FURTHER STEP FOR EFFICIENT CORRECTIONS OF INCONSISTENT PROBABILISTIC DATA SETS

Marco Baioletti[1] and Andrea Capotorti[1]

[1]Department of "Matematica e Informatica", University of Perugia - Italy
{*marco.baioletti,andrea.capotorti*} *@unipg.it*

Abstract

Partial conditional probability assessments are having renewed attention and one of the more compelling need associated with them is the of merging several sources of information. We focus here on the consequent mandatory task of correcting inconsistent probabilistic databases. Since probabilistic satisfiability problems (PSAT) has mainly suffered of space complexity in their original formulations, we propose an efficient method for correcting incoherent (i.e. inconsistent) conditional probability assessments. This method is based on $L1$ distance minimization and Mixed Integer Programming (MIP) procedures, taking into the right consideration the compulsory need and benefits of dealing with different "zero layers". Through a simple prototypical example, we illustrate the feasibility and the peculiarities of the proposed procedure

1 Introduction

Partial conditional probability assessments are having renewed attention since they appear as proper tools for managing "open worlds" and "on-demand" association rules in the Big Data era we are living in.

In fact, since the huge volume and heterogeneity of the data and the need to perform not previously designed analyses (e.g. in data-lake architectures), instead of a full and complex model, several smaller and specifically tailored tools are increasingly needed. And one of the more compelling need is that of merging several sources of information, e.g. as stated in (National Academies of Sciences Engineering and Medicine, 2017) that several agencies are currently investigating nonsurvey data sources to supplement or replace data from probability surveys. These investigations share common features, in which the information from the different sources needs to be evaluated and combined.

Fusion and merging has an associated risk of leading to an inconsistent information system. Moreover, whenever a part of a probabilistic evaluation needs to be revised, e.g. in dynamical models, inconsistencies can naturally appears and consequently a correction is compulsory.

As well outlined in Benferhat et al. (1998), the way data fusion problem is tackled depends on the way information is represented. Since one of the most familiar and adopted measure of uncertainty is probability, consequently we focus here on the even more mandatory task of correcting inconsistent probabilistic databases (see, e.g., Lian et al. (2010)).

The choice of correcting probability values reflects the willingness to maintain the probabilistic nature of the different sources of information. In fact it is required that the fusion would preserve the expressive power of probability framework. Of course, a change of the uncertainty management paradigm could be possible by adopting more general degrees of belief to deal with ill-posed sentences, like e.g. Belief functions, Fuzzy Logic, possibility measures or capacities (there is a vast literature on this, see among the others Bacchus et al. (1996); Benferhat et al. (1995); Benferhat and Sossai (2006); Bosc and Pivert (2013); Castro et al. (1994); Dubois et al. (2016); Miranda et al. (2021)) but this would be a strong intervention on the information representation, with a possible loss in expressiveness (as defined in Dubois et al. (2016)), especially if the fusion process is performed by a "third party" with respect the original sources. Hence we describe a way to proceed when the fusion process is intended to follow probability rules, and specifically those related to conditional probability assessments (see e.g. Biazzo and Gilio (2005); Coletti and Scozzafava (2001, 2002)).

Since probabilistic satisfiability problems (PSAT) has mainly suffered of space complexity in their original formulations, in the last decade we have proposed (see e.g. Baioletti and Capotorti (2015, 2018, 2020)) an efficient method for correcting incoherent (i.e. inconsistent) probability assessments. This method is based on $L1$ distance minimization and Mixed Integer Programming (MIP) procedures that can be designed to implement them, in line with what has been done in Cozman and Fargonidi Ianni (2013, 2015) where such technique was introduced for checking the coherence of a probabilistic assignment.

The aforementioned contributions mainly dealt with unconditional assessments, while correction of conditional probabilistic assessments was simply sketched. We are able now to propose an effective MIP implementation also for the conditional case, taking into the right consideration the compulsory need and benefits of dealing with different zero layers (see e.g. Coletti and Scozzafava (1997)).

Through a simple prototypical example we will illustrate the effective need of the iteration on different zero layers, while a detailed description of the MIP implementation will show the feasibility of the proposed procedure.

2 Conditional assessments and their correction

2.1 Conditional probability assessment

Definition 1 *A* ***conditional probability assessment*** *is a quadruple*

$$\pi = (\mathcal{V}, \mathcal{U}|\mathcal{H}, p, \mathcal{C}) \tag{1}$$

where:

- $\mathcal{V} = \{X_1, \ldots, X_k, H_1, \ldots, H_k\}$ *is a finite set of propositional variables, with* X_i*'s representing any potential event of interest and* H_i*'s their associated scenarios;*
- $\mathcal{U}|\mathcal{H}$ *is a set of* $n \leq k$ *effective conditional events* $X_i|H_i$*'s taken into consideration;*
- $p : \mathcal{U}|\mathcal{H} \rightarrow [0, 1]^n$ *is a vector which assigns a "potential" probability value* p_i*,* $i = 1, \ldots, n$*, to each conditional event* $X_i|H_i$*;*
- $\mathcal{C}$ *is a finite set of* ***logical constraints*** *which lie among all the variables in* $\mathcal{V}$*.*

Usually, probability values p are associated to the elements in $\mathcal{U}|\mathcal{H}$ and are assessed on the base of data or expert evaluations, but logical constraints $\mathcal{C}$ can be written in terms of all the potential events in $\mathcal{V}$, permitting to extend an initial assessment to a larger domain without redefining the whole model.

2.1.1 A focus on the logical constraints

Note that the constraints in $\mathcal{C}$ are written with the usual logical notation, where $\neg$, $\wedge$ and $\vee$ denote the negation, disjunction and conjunction connectives, respectively; $\Longrightarrow$ the material implication; $=$ the logical equivalence; $\top$ and $\perp$ the universal tautology and contradiction (sure and impossible events), respectively.

These constraints can be used to represent any kind of compound event, for instance that an event is the conjunction of other two events, or denote the implications or incompatibilities among the elements of $\mathcal{V}$.

Usually some possible forms of logical constraints are: $\phi = \psi$, $\phi \Longrightarrow \psi$ and $\phi = \perp$, where ψ and ϕ are boolean expressions involving the variables of $\mathcal{V}$. Without loss of generality, $\mathcal{C}$ can be expressed (explicitly or through automatized procedures) in conjunctive normal form (CNF), this will help in the implementation part of the correction procedure. Hence

$$\mathcal{C} = \{c_1, \ldots, c_m\} \tag{2}$$

where each element c_i of $\mathcal{C}$ is a disjunctive clause, i.e. it can be written as disjunction of literals formed with variables in $\mathcal{V}$

$$c_i = \left(\bigvee_{\gamma \in \Gamma_i} X_\gamma \right) \vee \left(\bigvee_{\lambda \in \Lambda_i} \neg X_\lambda \right) \tag{3}$$

for some $\Gamma_i, \Lambda_i \subseteq \{1, \ldots, n\}$. For example, the constraint $X_i \Longrightarrow X_j$ is expressed in $\mathcal{C}$ by the clause $\neg X_i \vee X_j$.

Since we will require that all the logical constraint present in $\mathcal{C}$ must be satisfied, $\mathcal{C}$ can be seen as the conjunction of $c_1, \ldots, c_m$, with m the number of disjunctive clauses present in $\mathcal{C}$.

In the sequel, to avoid a foolish solution (i.e. a probability mass function concentrated on $\bigwedge_{i=1}^n \neg H_i$)for the correction procedure, the constraint c_{m+1}

$$\bigvee_{i=1}^{n} H_i = \top \tag{4}$$

will be always added to $\mathcal{C}$.

Such logical constraints $\mathcal{C}$ are crucial to coerce the possible combinations of truth values among the events in $\mathcal{V}$. For this, we recall some basic definition:

Definition 2 *A* ***truth assignment*** *on $\mathcal{V}$ is a function $\alpha : \mathcal{V} \to \{0,1\}$. We denote by $2^{\mathcal{V}}$ the set of all truth assignments. We denote with $\alpha \models \phi$ the fact that the assignment $\alpha \in 2^{\mathcal{V}}$ satisfies a boolean expression ϕ (which means that replacing each variable x appearing in ϕ with the corresponding truth value $\alpha(x)$, the expression ϕ evaluates to 1).*

Definition 3 *A truth assignment $\alpha \in 2^{\mathcal{V}}$ is called* ***atom*** *for a probability assessment $\pi = (\mathcal{V}, \mathcal{U}|\mathcal{H}, p, \mathcal{C})$ if α satisfies all the logical constraints $c_i \in \mathcal{C}$, $i = 1, \ldots, m+1$.*

Definition 4 *The* ***the Boolean algebra spanned by*** *$\mathcal{V}$ and satisfying $\mathcal{C}$, denoted by $\mathcal{A}_{\mathcal{V}}$, is the power set of all the atoms of a probability assessment $\pi = (\mathcal{V}, \mathcal{U}|\mathcal{H}, p, \mathcal{C})$.*

Moreover, we will denote with $\mathcal{A}^0_{\mathcal{V}} = \mathcal{A}_{\mathcal{V}} \setminus \{\bot\}$

With these notions we can recall the consistency notion in the next section.

2.2 Coherence

As already mentioned, we focus our attention on inconsistent assessments π. Consistency for partial assessments can be reduced to the compatibility with a well established mathematical model. For conditional probabilities the reference models are the so called *full conditional probabilities*, as introduced by Dubins and in line also with De Finetti, Krauss and Rényi thoughts (for a detailed exposition on this subject refer to Coletti and Scozzafava (2002)).

Full conditional probabilities are characterized by the following set of axioms:

Definition 5 *Given a Boolean algebra $\mathcal{B}$, a* ***full conditional probability*** *on $\mathcal{B} \times \mathcal{B}^0$ ($\mathcal{B}^0 = \mathcal{B} \setminus \{\bot\}$) is a function $P : \mathcal{B} \times \mathcal{B}^0 \to [0,1]$ such that*
(i) $P(\cdot|H)$ is a finitely additive probability on $\mathcal{B}$ for any given H in $\mathcal{B}^0$;
(ii) $P(H|H) = 1$ for all $H \in \mathcal{B}^0$;
(iii) $P(A|C) = P(A|B)P(B|C)$ for every $A \in \mathcal{B}$, $B, C \in \mathcal{B}^0$, with $A \implies B \implies C$.

Note that, whenever (i) and (ii) are satisfied, condition (iii) is equivalent to
(iii') $P(AB|C) = P(B|C)P(A|BC)$ for every $A, B \in \mathcal{B}$, $C \in \mathcal{B}$, $BC \in \mathcal{B}^0$.

Consequently we have:

Definition 6 *A conditional probability assessment $\pi = (\mathcal{V}, \mathcal{U}|\mathcal{H}, p, \mathcal{C})$ as in Def.1 is said to be* ***coherent*** *if there exists a* full conditional probability *P as in Def.5 defined on $\mathcal{A}_{\mathcal{V}} \times \mathcal{A}^0_{\mathcal{V}}$ which agrees with p on $\mathcal{U}|\mathcal{H}$.*

Conditional assessments π given on arbitrary domains $\mathcal{U}|\mathcal{H}$ (for these such assessments are sometimes called also *partial*) have been fully investigated, both semantically (in conditional betting schemes) and syntactically (refer again to Coletti and Scozzafava (2002) for further details). We deal with them only syntactically and for this aim we recall the following characterization theorem given in Coletti and Scozzafava (2002) (we report it in an equivalent formulation adapted to the formalism used in this paper)

Theorem 1 *Let π be a conditional probability assessment as in Def.1. The following propositions are equivalent:*

- *π is a coherent conditional probability as in Def.6;*
- *there exists at least one finite class of unconditional probabilities $\{P_0, P_1, \dots, P_k\}$ such that:*
  1. *P_0 is defined over $\mathcal{A}_{\mathcal{V}}$, while for $\omega = 1, \dots, k$ the probability P_ω is defined over $\mathcal{A}_\omega$, i.e. the algebra generated by the atoms $\alpha_r \in \mathcal{A}_{\mathcal{V}}$ such that $P_{\omega-1}(\alpha_r) = 0$;*
  2. *for all $X_i|H_i \in \mathcal{U}|\mathcal{H}$ there exists a unique ω such that $P_\omega(H_i) > 0$ and*

$$p(X_i|H_i) = \frac{\sum_{\alpha_r \models X_i \wedge H_i} P_\omega(\alpha_r)}{\sum_{\alpha_r \models H_i} P_\omega(\alpha_r)}. \tag{5}$$

The different ω's represent different "zero layers" and to each H_i only one layer is associated.

Note that, in our models, conditional probabilities are assessed *directly* on conditional events $X_i|H_i$ and not as ratios of unconditional probabilities $P(X_i \wedge H_i)$ and $P(H_i)$ as in Kolmogorovian approach. This is possible because conditional events can be managed as "unitaries" logical entities, as well explained in Coletti and Scozzafava (2002) (we do not enter into such aspect in more details because it is out of the aim of this paper). Theorem 1 states that the representation by ratios is still valid, but with a class of suitable unconditional probabilities $\{P_0, P_1, \dots, P_k\}$ instead of a unique one.

Operationally speaking, coherence is noting more than specific constraints satisfactions on different "unexpectedeness" scenarios (the so called "zero layers") expressed through (5). Moreover, if a probability assessment is coherent, there exists a **sparse probability distribution** μ_ω for each "zero layer" (for more details in the unconditional case refer, e.g., to Jaumard et al. (1991)); namely, each P_ω is strictly positive on at most $n_\omega + 1$ atoms, denoted by

$$\alpha_\omega^{(1)}, \dots, \alpha_\omega^{(n_\omega+1)} \tag{6}$$

with $n_\omega = n$ for $\omega = 0$, otherwise it is the number of H_i s.t. $P_{\omega-1}(H_i) = 0$.

2.3 L1 correction

When a probability assessment $\pi = (\mathcal{V}, \mathcal{U}|\mathcal{H}, p, \mathcal{C})$ is not coherent, then it is possible to "correct" it in order to obtain a coherent probability assessment π' which is as close as possible to π, according to a distance or a pseudo-distance function between probability assessments. In this paper, as in our previous contributions, we focus on the distance between only the numerical parts p and p' of the assessments, all the other components $\mathcal{V}$, $\mathcal{U}|\mathcal{H}$ and $\mathcal{C}$ being the same.

One of the simpler distance between two numerical evaluations p and p' is the L_1 distance defined as

$$d_1(p, p') = \sum_{i=1}^{n} |p(X_i|H_i) - p'(X_i|H_i)| \quad (7)$$

and we denote by $\mathcal{C}(\pi)$ the sets of all the $L1$-corrections of π.

This distance, even being less "informative" with respect others (e.g. those based on entropy), has three important properties:

- the minimization of such displacements $|p(X_i|H_i) - p'(X_i|H_i)|$ obeys to the basic **principle of minimal change** in a numerical uncertainty setting;
- L_1 distance between probability distributions has important and nice **invariance properties** under strictly monotone transformations of random variables (Devroye and Gyorfi, 1985);
- the minimization of L_1 distance can be solved by MIP programming representing a clear **computational advantage** that is the topic of the present contribution.

2.4 A prototypical example

Example 1 *Let*

- $\mathcal{V} = \{X_1, \ldots, X_6, H_1, \ldots, H_6\}$
- $\mathcal{U}|\mathcal{H} = \{X_1|H_1, \ldots, X_6|H_6\}$
- $p = \{p(X_1|H_1) = 0.5, p(X_2|H_2) = 0.3, p(X_3|H_3) = 0.4, p(X_4|H_4) = 0.7, p(X_5|H_5) = 0.1, p(X_6|H_6) = 0.8\}$
- $\mathcal{C} = \{H_1 = H_2, H_2 = H_3, X_3 = (X_1 \wedge \neg(X_2)) \vee (\neg(X_1) \wedge X_2), X_1 \wedge X_2 \wedge H_4 = \bot, X_4 = X_1 \wedge H_4, X_1 \vee X_2 \vee H_4 \implies H_1, X_6 = X_4, X_5 \implies H_5, H_5 \wedge (X_1 \vee X_2 \vee H_4) = \bot, H_4 \implies H_6\}$

that turns out to be incoherent, since it should be $p(X_6|H_6) \leq p(X_4|H_4)$ *while we have* $p(X_6|H_6) = 0.8 > 0.7 = p(X_4|H_4)$.

Anyhow, this inconsistency can be detected only at the third layer, i.e. with $\omega = 2$, *since there exist a unconditional probability* P_0, *with* $P_0(H_5) = 1$ *while* $P_0(H_j) = 0$ *for* $j \in \{1, 2, 3, 4, 6\}$, *and an other one* P_1, *with* $P_1(H_1) > 0$, $P_1(H_2) > 0$ *and* $P_1(H_3) > 0$ *while* $P_1(H_4) = P_1(H_6) = 0$, *that permit to obtain as ratios* $p(X_5|H_5)$ *and* $\{p(X_1|H_1), p(X_2|H_2), p(X_3|H_3)\}$, *respectively.*

Hence there are infinite corrections $p' \in \mathcal{C}(\pi)$ *with minimal* $L1$ *distance of* $d_1(p, p') = 0.1$ *and they can be expressed through* $p'(X_6|H_6) = 0.8 - x$ *and* $p'(X_4|H_4) = 0.7 + x$, $x \in [0.05, 0.1]$, *while* $p'(X_j|H_j) = p(X_j|H_j)$, $j \in \{1, 2, 3, 5\}$.

If we add the further objective of minimizing the number of corrections, i.e. optimizing the size of the so called "minimal explanations" (O'Sullivan et al., 2007), we have the unique solution obtained by taking $x = 0.1$.

3 Correction with MIP

In this section we will describe an algorithm which finds a L1 correction for a conditional probability assessment π, as defined in the previous Section, by solving a sequence of Mixed Integer Programming (MIP) problems (Wolsey, 2008). Each MIP problem correspond to correct the probability values on a ”zero layer”, as described in Theorem 1.

A MIP problem is the problem of minimizing (or maximizing) a linear objective function $f(x_1, \ldots, x_r, k_1, \ldots, k_s)$, where $x_1, \ldots, x_r$ are non negative real variables, and $k_1, \ldots, k_s$ are non negative integer variables, which are constrained by a finite set of linear inequalities.

Modern MIP solvers are quite efficient, especially commercial solvers, in fact they are able to solve large instances of MIP in a reasonable amount of time.

3.1 MIP formulation

We start our description by explaining how a single MIP program is designed. We suppose that the correction operates on the layer ω with n_ω conditional events. We denote by $Y_1, \ldots, Y_{2n_\omega}$ all the propositional variables needed to describe the conditional events in the layer, i.e., $Y_1 = E_1, Y_2 = H_1, \ldots, Y_{2n_\omega - 1} = E_n, Y_{2n_\omega} = H_{n_\omega}$.

The formulation of the problem of minimizing the distance (7) as MIP employs one set of binary variables and 5 sets of real variables.

The first set comprises the $2n_\omega \cdot (n_\omega + 1)$ binary variables a_{ij} which corresponds to the unknown value assigned by each of the $n_\omega + 1$ atoms of the sparse solution (6) to all the propositional variables $\alpha_\omega^{(j)}(Y_i)$, for $i = 1, \ldots, 2n_\omega$ and $j = 1, \ldots, n_\omega + 1$.

The second set is composed by $n_\omega + 1$ real non negative variables q_j, for $j = 1, \ldots, n_\omega + 1$, which correspond to the unknown probability values $P_\omega(\alpha_\omega^{(j)})$.

The third set is composed by $n_\omega \cdot (n_\omega + 1)$ real non negative variables b_{ij}, for $i = 1, \ldots, n_\omega$ and $j = 1, \ldots, n_\omega + 1$, which are used to express the product $q_j \cdot a_{2i-1,j} \cdot a_{2i,j}$ Using b_{ij}, it is possible to maintain the linearity of the problem, without using explicit multiplications.

Similarly, the fourth set is composed by the $n_\omega \cdot (n_\omega + 1)$ real non negative variables b'_{ij}, for $i = 1, \ldots, n_\omega$ and $j = 1, \ldots, n_\omega + 1$, which are used to express the product $q_j \cdot a_{2i,j}$.

Finally, the fifth and the sixth sets are composed by the n_ω real non negative variables r_i and s_i, which are used to correct the corresponding probability values $P(Y_{2i-1}|Y_{2i}) = P(E_i|H_i)$. The value of r_i (respectively, s_i) is used to increase (decrease) the value $P(Y_{2i-1}|Y_{2i})$.

The objective function to be minimized in the MIP is

$$\sum_{i=1}^{n_\omega} (r_i + s_i)$$

The linear constraints are divided in 5 groups.

In the first group, each probability value $P(Y_{2i-1}|Y_{2i})$ is corrected with the equation

$$\sum_{j=1}^{n_\omega+1} b_{ij} = p(X_i|H_i) \cdot \sum_{j=1}^{n_\omega+1} b'_{ij} + r_i - s_i$$

for $i = 1, \dots, n_\omega$.

The second group is composed by the normalization constraint

$$\sum_{j=1}^{n_\omega+1} q_j = 1.$$

The third group enforces the relations $b_{ij} = q_j \cdot a_{2i-1,j} \cdot a_{2i,j}$ by posing the following linear constraints

$$0 \leq b_{ij} \leq a_{2i-1,j}$$

$$b_{ij} \leq a_{2i,j}$$

$$a_{2i-1,j} + a_{2i,j} - 2 + q_j \leq b_{ij} \leq q_j$$

for $i = 1, \dots, n_\omega$ and $j = 1, \dots, n_\omega + 1$.

Analogously, the fourth group is related to the relation $b'_{ij} = q_j \cdot a_{2i,j}$, by adding the linear constraints

$$0 \leq b'_{ij} \leq a_{2i,j}$$

$$a_{2i,j} - 1 + q_j \leq b'_{ij} \leq q_j$$

for $i = 1, \dots, n_\omega$ and $j = 1, \dots, n_\omega + 1$.

The additional clause c_{m+1}, that corresponds to the logical constraint (4), is added to the CNF clauses.

Finally, the fifth group imposes that each assignment $\alpha_\omega^{(j)}$ is an atom. For each clause c_i, let Γ_i and Λ_i the sets of variables appearing in c_i affirmed and negated, respectively, as stated in (3). Then, the constraints

$$\sum_{h \in \Gamma_i} a_{hj} + \sum_{l \in \Lambda_i} (1 - a_{lj}) \geq 1$$

are added, for each $j = 1, \dots, n_\omega + 1$ and for $i = 1, \dots, m + 1$.

The solution of the MIP problem provides the following results:

- the value of the objective function is the L1 distance (7) between the initial assessment p and a corrected one p', both restricted to $\{E_1|H_1, \dots, E_{n_\omega}|H_{n_\omega}\}$;
- for all $i = 1, \dots, n_\omega$ $P_\omega(H_i)$ can be computed as

$$P_\omega(H_i) = \sum_{j=1}^{n_\omega+1} b'_{ij}$$

;

- for all the events H_i such that $P_\omega(H_i) > 0$, the corrected value $p'(X_i|H_i)$ can be computed as

$$p'(X_i|H_i) = p(X_i|H_i) + \frac{r_i - s_i}{P_\omega(H_i)} \tag{8}$$

Overall, the MIP problem has $2n_\omega(n_\omega + 1)$ binary variables and $2n_\omega^2 + 5n_\omega + 1$ real variables, and contains $9n_\omega^2 + n_\omega(9 + 2n_\omega + m + 1) + 1$ linear constraints. Hence, the MIP formulation is able to encode the correction problem for a single layer with a polynomial number of variables and constraints, with respect to the size of the assessment.

It is important to notice that the other approaches to the checking of coherence does not have the nice property of polynomial space requirements. In fact, the linear system used to check the coherence has an exponential number of columns (in the worst case) and to overcome this feature, iterative techniques, such as column generation, must be adopted to reduce the space requirement.

Another important aspect of using MIP formulation is the possibility to exploit fast implementation of MIP solvers, without recurring to an ad-hoc algorithm. The MIP formulation is however strictly related to the form of distance used to express the correction. In fact, other more popular choices for the distance (e.g., the L2 or "Kullback-Leibler like" distances) will require a nonlinear (quadratic or logarithmic) programming solver, which is much more demanding from the computational point of view.

3.2 Iterative algorithm

The algorithm for correcting the entire assessment is described in the algorithm 1.

```
input  : the probability assessment
output: the corrected values
I ← {1, . . . , n} ;
ω ← 0;
while I ≠ ∅ do
    create the MIP problem for the assessment restricted to the events in I;
    solve it and extract the set of variables b′, r, and s;
    compute P_ω(H_i) for all i ∈ I;
    J ← {i ∈ I : P_ω(H_i) > 0};
    for i ∈ J do
        store the corrected value p(X_i|H_i) + (r_i − s_i)/P_ω(H_i);
    end
    I ← I \ J;
    ω ← ω + 1;
end
```

Algorithm 1: Iterative Algorithm for correcting a conditional probability assessment π

The algorithm initially tries to correct the entire assessment by creating the MIP problem and by solving it with a MIP solver. Then, all the values of $P_\omega(H_i)$ are computed by using the variables b' extracted from the solution produced. For all the events H_i such that $P_\omega(H_i)$ is 0, no correction can be found because in the corresponding equation

$$\sum_{j=1}^{n_\omega+1} b_{ij} = p(X_i|H_i) \cdot \sum_{j=1}^{n_\omega+1} b'_{ij} + r_i - s_i$$

also the value of $P_\omega(E_i \wedge H_i) = \sum_{j=1}^{n_\omega+1} b_{ij}$ is 0 and then this equation is trivially satisfied by setting $r_i = s_i = 0$.

On the other hand, if $P_\omega(H_i) > 0$, then the correction of $p(X_i|H_i)$ is computed with Equation (8).

The algorithm continues on the next layer by restricting the correction only on those conditional events $X_i|H_i$ such that $P_\omega(H_i) = 0$. Moreover, the additional clause c_{m+1} expressed as in (4) is updated according to the new set of indices I.

At the end, the algorithm has found the corrected value for the probability of each conditional event.

It is important to understand that, in general, there can be infinite corrections of a probability assessment: the choice of which correction is returned depends by the algorithm on the MIP solver used to solve the MIP problem. In any case, the algorithm is always able to find a valid correction. However, it is possible to select corrections which satisfy particular conditions, by modifying the objective function of the MIP problem.

A prototypical implementation of the correction algorithm has been realized in Python (using CPLEX^{TM} as MIP solver) and it is freely available at `https://github.com/mbaioletti/CorrCondProb`.

4 Conclusions and future work

With this contribution we propose optimization tasks for the fusion and correction of inconsistent conditional probability assessments through mixed-integer programming (MIP) techniques.

One of the peculiarity of the proposal is an effective MIP implementation specifically tailored for the conditional case, taking into the right consideration the compulsory need and benefits of dealing with different unexpectedness of different scenarios (the so-called "zero layers).

The sets of coherent conditional assessments are quite subtle mathematical entities, e.g. they lack, in general, of convexity (as stated e.g. in (Gilio, 1999)). Hence, the correction procedure we have illustrated here could produce not-unique and tangled solutions (e.g. union of disconnected parts with some isolated point...) and so, either a further refinement step to propose a unique result, or an extension to other uncertainty measures (e.g. conditional belief functions, possibilities or capacities) are needed.

Acknowledgments

This research has been founded by the the University of Perugia's DMI Department reasearch funds "Fondo di funzionamento per la ricerca dipartimentale" years 2019-2022 of both authors.

References

F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. From statistical knowledge bases to degrees of belief. *Artificial Intelligence*, 87(1—2):75–143, 1996. doi: 0.1016/S0004-3702(96)00003-3.

M. Baioletti and A. Capotorti. Efficient l1-based probability assessments correction: Algorithms and applications to belief merging and revision. In *ISIPTA'15 Proc. of the 9th International Symposium on Imprecise Probability: Theories and Applications. Pescara (IT)*, pages 37–46. ARACNE, 2015.

M. Baioletti and A. Capotorti. A $l1$ based probabilistic merging algorithm and its application to statistical matching. *Applied Intelligence*, 49:1–13, 2018. doi: 10.1007/s10489-018-1233-z.

M. Baioletti and A. Capotorti. A l1 minimization optimal corrective explanation procedure for probabilistic databases. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 80–92. 2020.

S. Benferhat and C. Sossai. Reasoning with multiple-source information in a possibilistic logic framework. *Information Fusion*, 7:80–96, 2006. doi: 10.1016/j.inffus.2005.01.006.

S. Benferhat, D. Dubois, and H. Prade. How to infer from inconsistent beliefs without revising? In *Proc. of the 14th Inter. Joint Conf. on Artificial Intelligence (IJCAI'95), Montréal, Canada, Aug. 20–25*, pages 1449—-1455, 1995.

S. Benferhat, D. Dubois, and H. Prade. From semantic to syntactic approaches to information combination in possibilistic logic. In B. Bouchon-Meunier, editor, *Aggregation and Fusion of Imperfect Information*, pages 141–161. Physica-Verlag HD, 1998. doi: 10.1007/978-3-7908-1889-5_9.

V. Biazzo and A. Gilio. Some theoretical properties of conditional probability assessments. *LNAI (Lecture Notes in Computer Science)*, 3571:775–787, 2005.

P. Bosc and O. Pivert. Querying possibilistic databases: Three interpretations. In R. R. Yager, A. M. Abbasov, M. Z. Reformat, and S. N. Shahbazova, editors, *Soft Computing: State of the Art Theory and Novel Applications*, pages 161–176. Springer, Berlin, Heidelberg, 2013. doi: 10.1007/978-3-642-34922-5_12.

J. Castro, F. Herrera, and J. Verdegay. Knowledge-based systems and fuzzy boolean programming. *Intelligent Systems*, 9(2):211–225, 1994. doi: 10.1002/int.4550090203.

G. Coletti and R. Scozzafava. Exploiting zero probabilities. In *Proc. EUFIT '97*, ELITE Foundation, pages 1499–1503, Aachen, 1997.

G. Coletti and R. Scozzafava. From conditional events to conditional measures: A new axiomatic approach. *Annals of Mathematics and Artificial Intelligence*, 32:373—392, 2001.

G. Coletti and R. Scozzafava. *Probabilistic Logic in a Coherent Setting.* Trends in Logic. Kluwer, Dordrecht, 2002. doi: 10.1007/978-94-010-0474-9.

F. Cozman and L. Fargonidi Ianni. Probabilistic satisfiability and coherence checking through integer programming. *Lecture Notes in Computer Science*, 7958:145–156, 2013. doi: 10.1007/978-3-642-39091-3_13.

F. Cozman and L. Fargonidi Ianni. Probabilistic satisfiability and coherence checking through integer programming. *International Journal of Approximate Reasoning*, 58: 57–70, 2015. doi: 10.1016/j.ijar.2014.09.002.

L. Devroye and L. Gyorfi. *Nonparametric Density Estimation: The L1 View.* Wiley Interscience Series in Discrete Mathematics. Wiley, 1985. ISBN 9780471816461. URL `https://books.google.co.il/books?id=ZVALbrjGpCoC`.

D. Dubois, W. Liu, J. Ma, and H. Prade. The basic principles of uncertain information fusion. an organised review of merging rules in different representation frameworks. *Information Fusion*, 32(Part A):12–39, 2016. doi: 10.1016/j.inffus.2016.02.006.

A. Gilio. Probabilistic relations among logically dependent conditional events. *Soft Computing*, 3:154–161, 1999.

B. Jaumard, P. Hansen, and M. P. de Aragão. Column generation methods for probabilistic logic. *ORSA Journal on Computing*, 3(2):135–148, 1991.

A. Lian, A. Chen, and A. Song. Consistent query answers in inconsistent probabilistic databases. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 303–314, Indianapolis, Indiana, 2010. doi: 10.1145/1807167.1807202.

E. Miranda, I. Montes, and P. Vicig. On the selection of an optimal outer approximation of a coherent lower probability. *Fuzzy Sets and Systems*, 424:1–36, 2021.

. National Academies of Sciences Engineering and Medicine. *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps.* The National Academies Press, Washington, DC, 2017. doi: 10.17226/24893.

B. O'Sullivan, A. Papadopoulos, B. Faltings, and P. Pu. Representative explanations for over-constrained problems. *Proceedings of the National Conference on Artificial Intelligence*, 1, 07 2007.

L. A. Wolsey. *Mixed Integer Programming*, pages 1–10. John Wiley & Sons, Ltd, 2008. ISBN 9780470050118. doi: https://doi.org/10.1002/9780470050118.ecse244.

Selfadhesivity in Gaussian conditional independence structures

Tobias Boege

Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig
tobias.boege@mis.mpg.de

Abstract

Selfadhesivity is a property of entropic polymatroids which can be formulated as gluability conditions of the polymatroid to an identical copy of itself along arbitrary restrictions and such that the two pieces are independent given the common restriction. We show that positive definite matrices satisfy this condition as well and examine consequences for Gaussian conditional independence structures. New axioms of Gaussian CI are obtained by applying selfadhesivity to the previously known axioms of structural semigraphoids and orientable gaussoids.

1 Introduction

In matroid theory, the term *amalgam* refers to a gluing of two matroids along a common restriction, similar to how four triangles can be glued together along edges to form the boundary of a tetrahedron. This concept is meaningful for conditional independence (CI) structures as well. The bridge from the geometric (matroid-theoretical) concept to probability theory (conditional independence) is built by Matúš (2007) who defines a special kind of amalgam, the *adhesive extension*, for polymatroids and proves that such extensions always exist for entropic polymatroids with a common restriction.

The purpose of this article is two-fold: First, it is to abstract further than polymatroids and to introduce a derived collection of amalgamation properties known as *selfadhesivity* for general conditional independence structures and apply the mechanism of selfadhesion to strengthen already known conditional independence inference rules. Second, this general treatment of selfadhesivity is driven by its applications to Gaussian instead of discrete CI inference. The main result, Theorem 1, shows that also in the Gaussian setting adhesive extensions (of covariance matrices) exist and are even unique. We use the non-trivial gluability conditions implied by this result to derive new axioms for Gaussian conditional independence structures.

2 Preliminaries

Gaussian conditional independence. Let N be a finite ground set indexing jointly distributed random variables $\xi = (\xi_i : i \in N)$. By convention, elements of N are denoted by $i, j, k, \ldots$ and subsets by $I, J, K, \ldots$. Elements are identified with singleton subsets of N and juxtaposition of subsets abbreviates set union. Thus, an expression such as iK is shorthand for $\{ i \} \cup K$ as a subset of N. The complement of $K \subseteq N$ is K^{c}.

We are mostly interested in Gaussian (i.e., multivariate normal) distributions. These distributions are specified by a small number of parameters, namely the mean vector $\mu \in \mathbb{R}^N$ and the covariance matrix $\Sigma \in \mathrm{PD}_N$. Throughout this article, "Gaussian" means "regular Gaussian", i.e., the covariance matrix is strictly positive definite. On the boundary of the PD-cone, for positive semidefinite covariance matrices, the CI theory is algebraically more complicated and valid inference properties for regular Gaussians can fail to be valid for singular ones; see Studený (2005), Section 2.3.6.

The following result summarizes basic facts from algebraic statistics relating subvectors of ξ and their (positive definite) covariance matrices. It can be found, for instance, in §2.4 of Sullivant (2018). For $\Sigma \in \mathrm{PD}_N$ and $I, J, K \subseteq N$, let $\Sigma_{I,J}$ denote the submatrix with rows indexed by I and columns by J. Submatrices of the form $\Sigma_K \coloneqq \Sigma_{K,K}$ are *principal*. Dual to a principal submatrix is its *Schur complement* $\Sigma^K \coloneqq \Sigma_{K^{\mathsf{c}}} - \Sigma_{K^{\mathsf{c}},K}\Sigma_K^{-1}\Sigma_{K,K^{\mathsf{c}}}$. Principal submatrices and Schur complements of positive definite matrices are also positive definite.

Theorem. *Let ξ be distributed according to the (regular) Gaussian distribution with mean $\mu \in \mathbb{R}^N$ and covariance $\Sigma \in \mathrm{PD}_N$. Let $K \subseteq N$.*

— *The marginal vector $\xi_K = (\xi_k : k \in K)$ is a regular Gaussian in $\mathbb{R}^K$ with mean vector μ_K and covariance Σ_K.*

— *Let $y \in \mathbb{R}^K$. The conditional $\xi_{K^{\mathsf{c}}} \mid \xi_K = y$ is a regular Gaussian in $\mathbb{R}^{K^{\mathsf{c}}}$ with mean vector $\mu_{K^{\mathsf{c}}} + \Sigma_{K^{\mathsf{c}},K}\Sigma_K^{-1}(y - \mu_K)$ and covariance Σ^K.*

— *Let a Gaussian over $N = IJ$ be given with covariance $\Sigma \in \mathrm{PD}_{IJ}$. Then the marginal independence $[\xi_I \perp\!\!\!\perp \xi_J]$ holds if and only if $\Sigma_{I,J} = 0$.*

The general CI statement $[\xi_I \perp\!\!\!\perp \xi_J \mid \xi_K]$ is the result of marginalizing ξ to IJK, conditioning on K and then checking for independence of I and J. The previous lemma implies the following algebraic CI criterion for regular Gaussians:

$$\begin{aligned}
[\xi_I \perp\!\!\!\perp \xi_J \mid \xi_K] \;&\Leftrightarrow\; \left(\Sigma_{IJ} - \Sigma_{IJ,K}\Sigma_K^{-1}\Sigma_{K,IJ}\right)_{I,J} = 0 \\
&\Leftrightarrow\; \Sigma_{I,J} - \Sigma_{I,K}\Sigma_K^{-1}\Sigma_{K,J} = 0 \\
&\Leftrightarrow\; \operatorname{rk}\Sigma_{IK,JK} = |K|. \qquad (\perp\!\!\!\perp)
\end{aligned}$$

The last equivalence follows from rank additivity of the Schur complement (see Zhang (2005)) and the observation that the $K \times K$ submatrix of Σ has full rank $|K|$ because it is a positive definite matrix. In particular, the truth of a conditional independence statement does not depend on the conditioning event and it does not depend on the mean μ. Hence we identify regular Gaussians with their covariance matrices $\Sigma \in \mathrm{PD}_N$.

The "$\geq$" part of the rank condition in ($\perp\!\!\!\perp$) always holds because the principal submatrix Σ_K has full rank $|K|$. Then this minimal rank of $\Sigma_{IK,JK}$ is attained if and only if all of its minors of size $|K|+1$ vanish. These minors correspond to CI statements of the form $[\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K]$ with $i \in I$ and $j \in J$. This proves the following "localization rule" for Gaussian conditional independence:

$$[\xi_I \perp\!\!\!\perp \xi_J \mid \xi_K] \;\Leftrightarrow\; \bigwedge_{i \in I, j \in J} [\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K]. \tag{L}$$

Rules of this form go back to Matúš (1992). A weaker localization rule holds for all semigraphoids, whereas the one presented above can be proved for compositional graphoids. In both cases, a general CI statement is reduced to a conjunction of *elementary CI statements* $[\xi_i \perp\!\!\!\perp \xi_j \mid \xi_K]$ about the independence of two singletons. We adopt the form $[I \perp\!\!\!\perp J \mid K]$ for CI statements $[\xi_I \perp\!\!\!\perp \xi_J \mid \xi_K]$ without the mention of a random vector. These symbols are treated as combinatorial objects and $\mathcal{A}_N \coloneqq \{\, [i \perp\!\!\!\perp j \mid K] : ij \in \binom{N}{2}, K \subseteq N \setminus ij \,\}$ is the set of all elementary CI statements. The *CI structure* of Σ is the set

$$[\![\Sigma]\!] \coloneqq \{\, [i \perp\!\!\!\perp j \mid K] \in \mathcal{A}_N : \det \Sigma_{iK,jK} = 0 \,\}\,.$$

The localization rule shows that $[\![\Sigma]\!]$ encodes the entire set of true CI statements for a Gaussian with covariance matrix Σ and with slight abuse of notation we employ statements such as $[I \perp\!\!\!\perp J \mid K] \in [\![\Sigma]\!]$. It is important to note in this context that we treat only *pure* CI statements, i.e., $[I \perp\!\!\!\perp J \mid K]$ where I, J, K are pairwise disjoint. Any general CI statement with overlaps between the three sets decomposes, analogously to the localization rule, into a conjunction of pure CI statements and functional dependence statements. For a regular Gaussian, functional dependences are always false, so this is no restriction in generality. In particular, the general statement $[N \perp\!\!\!\perp M \mid L]$, which frequently appears later, is equivalent to $[(N \setminus L) \perp\!\!\!\perp (M \setminus L) \mid L]$ and hence is pure provided that $L \supseteq N \cap M$.

Polymatroids and selfadhesivity. A *polymatroid* over the finite ground set N is a function $h : 2^N \to \mathbb{R}$ assigning to every subset $K \subseteq N$ a real number, such that h is

normalized: $h(\emptyset) = 0$,

isotone: $h(I) \leq h(J)$ for $I \subseteq J$,

submodular: $h(I) + h(J) \geq h(I \cup J) + h(I \cap J)$.

With the linear functional $\triangle(I, J|K) \cdot h \coloneqq h(IK) + h(JK) - h(IJK) - h(K)$, submodularity on pairwise disjoint I, J, K can be restated as $\triangle(I, J|K) \cdot h \geq 0$. If h_ξ is the entropy vector of a discrete random vector ξ, i.e., $h(K)$ is the Shannon entropy of the marginal vector ξ_K, then it is a polymatroid and the quantity $\triangle(I, J|K) \cdot h_\xi$ is known as the *conditional mutual information* $I(\xi_I; \xi_J | \xi_K)$. Its vanishing is equivalent to the conditional independence $[\xi_I \perp\!\!\!\perp \xi_J \mid \xi_K]$. Hence we may define the CI structure of a polymatroid as $[\![h]\!] \coloneqq \{\, [i \perp\!\!\!\perp j \mid K] \in \mathcal{A}_N : \triangle(ij|K) \cdot h = 0 \,\}$. These structures are called *(elementary) semimatroids* in Matúš (1994) and (equivalently, but based on properties of multiinformation instead of entropy vectors) *structural semigraphoids* in Studený (1994).

Again, per Matúš (1994) a localization rule holds for them which we use to interpret the containment of non-elementary CI statements:

$$[I \perp\!\!\!\perp J \mid K] \in [\![h]\!] \;\Leftrightarrow \bigwedge_{\substack{i \in I, j \in J, \\ K \subseteq L \subseteq IJK \setminus ij}} [i \perp\!\!\!\perp j \mid L] \in [\![h]\!]. \tag{L'}$$

Matúš (2007) introduced the notion of adhesive extensions and selfadhesive polymatroids to mimic a curious amalgamation property of entropy vectors. The underlying construction is the *Copy lemma* of Zhang and Yeung (1998), also known as the *conditional product*; see Studený (2005), Section 2.3.3. Let g and h be two polymatroids on ground sets N and M, respectively, and suppose that their restrictions $g|_L$ and $h|_L$ to $L = N \cap M$ coincide. A polymatroid f on NM is an *adhesive extension* of g and h if:

— $f|_N = g$ and $f|_M = h$,

— $[N \perp\!\!\!\perp M \mid L] \in [\![f]\!]$.

Since $L \subseteq N$ and $L \subseteq M$, the statement $[N \perp\!\!\!\perp M \mid L]$ is naturally equivalent to the pure CI statement $[N' \perp\!\!\!\perp M' \mid L]$ with $N' = N \setminus L$ and $M' = M \setminus L$. In polymatroid terms, N and M are said to form a *modular pair* in f if this CI statement holds.

Next, suppose that we have only one polymatroid h on ground set N and fix $L \subseteq N$. An *L-copy* of N is a finite set M with $|M| = |N|$ and $M \cap N = L$. We fix a bijection $\pi : N \to M$ which preserves L pointwise. h is a *selfadhesive polymatroid at L* if there exists an adhesive extension of h and its induced copy $\pi(h)$ over their common restriction to L. The polymatroid is *selfadhesive* if it is selfadhesive at every $L \subseteq N$. The fundamental result of Matúš (2007) is:

Theorem. *Any two of the restrictions of an entropic polymatroid have an entropic adhesive extension. In particular, entropy vectors are selfadhesive.*

Furthermore, the set of polymatroids on a 4-element ground set which are selfadhesive forms a rational, polyhedral cone in $\mathbb{R}^{16}$. This cone is characterized (in addition to the polymatroid properties) by the validity of the Zhang-Yeung inequalities (see Remark 1). In this sense, selfadhesivity is a reformulation of the Zhang-Yeung inequalities using only the notions of restriction and conditional independence. Generalizations of adhesive extensions to multiple polymatroids lead to *book inequalities* from Csirmaz (2014).

3 Adhesive extensions of Gaussians

The analogous result for Gaussian covariance matrices is our main theorem:

Theorem 1. *Let $\Sigma \in \mathrm{PD}_N$ and $\Sigma' \in \mathrm{PD}_M$ be two covariance matrices with common restriction $\Sigma_L = \Sigma'_L$, where $L = N \cap M$. There exists a unique $\Phi \in \mathrm{PD}_{NM}$ such that:*

— *$\Phi_N = \Sigma$ and $\Phi_M = \Sigma'$,*

— *$[N \perp\!\!\!\perp M \mid L] \in [\![\Phi]\!]$.*

Proof. Let $N' = N \setminus L$, $M' = M \setminus L$. We use the following names for blocks of Σ and Σ':

$$\Sigma = \begin{array}{c} \\ {\scriptstyle L} \\ {\scriptstyle N'} \end{array} \begin{array}{c} \begin{array}{cc} {\scriptstyle L} & {\scriptstyle N'} \end{array} \\ \begin{pmatrix} X & A \\ A^{\mathsf{T}} & Y \end{pmatrix} \end{array}, \qquad \Sigma' = \begin{array}{c} \\ {\scriptstyle L} \\ {\scriptstyle M'} \end{array} \begin{array}{c} \begin{array}{cc} {\scriptstyle L} & {\scriptstyle M'} \end{array} \\ \begin{pmatrix} X & B \\ B^{\mathsf{T}} & Z \end{pmatrix} \end{array}.$$

Consider the matrix

$$\Phi = \begin{array}{c} \\ {\scriptstyle L} \\ {\scriptstyle N'} \\ {\scriptstyle M'} \end{array} \begin{array}{c} \begin{array}{ccc} {\scriptstyle L} & {\scriptstyle N'} & {\scriptstyle M'} \end{array} \\ \begin{pmatrix} X & A & B \\ A^{\mathsf{T}} & Y & \Lambda \\ B^{\mathsf{T}} & \Lambda^{\mathsf{T}} & Z \end{pmatrix} \end{array},$$

where Λ will be determined shortly. Its restrictions to N and M are clearly equal to Σ and Σ', respectively. By the rank additivity formula for Schur complements,

$$\operatorname{rk} \Phi_{N,M} = \operatorname{rk} \begin{pmatrix} X & B \\ A^{\mathsf{T}} & \Lambda \end{pmatrix} = \underbrace{\operatorname{rk} X}_{=|L|} + \operatorname{rk}(\Lambda - A^{\mathsf{T}} X^{-1} B),$$

and the rank requirement $\operatorname{rk} \Phi_{N,M} = |N \cap M| = |L|$, which is equivalent to $[N \perp\!\!\!\perp M \mid L]$, it is necessary that $\Lambda = A^{\mathsf{T}} X^{-1} B$. Thus, Φ is uniquely determined by Σ and Σ' via the two conditions in the theorem. To show positive definiteness, consider the transformation

$$P = \begin{array}{c} \\ {\scriptstyle L} \\ {\scriptstyle N'} \\ {\scriptstyle M'} \end{array} \begin{array}{c} \begin{array}{ccc} {\scriptstyle L} & {\scriptstyle N'} & {\scriptstyle M'} \end{array} \\ \begin{pmatrix} \mathbb{1} & -X^{-1}A & -X^{-1}B \\ 0 & \mathbb{1} & 0 \\ 0 & 0 & \mathbb{1} \end{pmatrix} \end{array}$$

of the bilinear form Φ:

$$P^{\mathsf{T}} \Phi P = \begin{pmatrix} X & 0 & 0 \\ 0 & Y - A^{\mathsf{T}} X^{-1} A & 0 \\ 0 & 0 & Z - B^{\mathsf{T}} X^{-1} B \end{pmatrix} = \begin{pmatrix} \Sigma_L & 0 & 0 \\ 0 & \Sigma^L & 0 \\ 0 & 0 & \Sigma'^L \end{pmatrix}.$$

The result is clearly positive definite and since P is invertible, this shows $\Phi \in \mathrm{PD}_{NM}$. $\square$

Remark 1. *Zhang and Yeung (1998) proved the first information inequality for the entropy region which is not a consequence of the Shannon inequalities (equivalently, the polymatroid properties). It can be expressed as the non-negativity of the functional*

$$\nabla(i,j|kl) \coloneqq \triangle(kl|i) + \triangle(kl|j) + \triangle(ij|) - \triangle(kl|) + \triangle(ik|l) + \triangle(il|k) + \triangle(kl|i).$$

Matúš (2007) characterized the selfadhesive polymatroids over a 4-element ground set as those polymatroids satisfying $\nabla(i,j|kl) \geq 0$ for all choices of i, j, k, l. As a corollary to Theorem 1 we obtain that the multiinformation vectors and hence the differential entropy vectors of Gaussian distributions satisfy the Zhang-Yeung inequalities. This is one half of the result proved by Lněnička (2003). However, that result also follows from the metatheorem of Chan (2003) since $\nabla(i,j|kl)$ is balanced.

In the theory of regular Gaussian conditional independence structures, it is natural to relax the positive definiteness assumption on Σ to that of *principal regularity*, i.e., all principal minors, instead of being positive, are required not to vanish. Principal regularity is the minimal technical condition which allows the formation of all Schur complements and the property is inherited by principal submatrices and Schur complements, hence enabling analogues of marginalization and conditioning over general fields instead of the field $\mathbb{R}$; see Boege (2021) for applications. However, the last step in the above proof requires positive definiteness and does not work for principally regular matrices:

Example 1. *Consider the following principally regular matrix over $N = ijkl$:*

$$\Gamma = \begin{array}{c} \\ i \\ j \\ k \\ l \end{array} \begin{array}{c} \begin{array}{cccc} i & j & k & l \end{array} \\ \left(\begin{array}{cc|cc} 1 & 0 & 7/8 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 7/8 & 0 & 1 & -\sqrt{1695}/64 \\ 0 & 0 & -\sqrt{1695}/64 & 1 \end{array}\right) \end{array}$$

and fix $L = ij$. By the proof of Theorem 1, the submatrix and rank conditions uniquely determine an adhesive extension of Γ with an L-copy of itself over the ground set $ijklk'l'$. This unique candidate matrix is

$$\begin{array}{c} \\ i \\ j \\ k \\ l \\ k' \\ l' \end{array} \begin{array}{c} \begin{array}{cccccc} i & j & k & l & k' & l' \end{array} \\ \left(\begin{array}{cc|cc|cc} 1 & 0 & 7/8 & 0 & 7/8 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 7/8 & 0 & 1 & -\sqrt{1695}/64 & 49/64 & 0 \\ 0 & 0 & -\sqrt{1695}/64 & 1 & 0 & 0 \\ \hline 7/8 & 0 & 49/64 & 0 & 1 & -\sqrt{1695}/64 \\ 0 & 0 & 0 & 0 & -\sqrt{1695}/64 & 1 \end{array}\right) \end{array}.$$

But this matrix is not principally regular, as the klk'-principal minor is zero.

4 Structural selfadhesivity

The existence of adhesive extensions and in particular selfadhesivity of positive definite matrices induces similar properties on their CI structures, since the conditions in Theorem 1 can be formulated using only the concepts of restriction and conditional independence. On the CI level, we sometimes use the term *structural selfadhesivity* to emphasize that it is generally a weaker notion than what is proved for covariance matrices above. Selfadhesivity can be used to strengthen known properties of CI structures: if it is known that all positive definite matrices have a certain distinguished property $\mathfrak{p}$, then the fact that Σ and any L-copy of it fit into an adhesive, positive definite extension obeying $\mathfrak{p}$ says more about the structure of Σ than $\mathfrak{p}$ alone. We begin by making precise the notion of a *property*:

Definition 1. *Let $\mathfrak{A}_N = 2^{\mathcal{A}_N}$ be the set of all CI structures over N. For $N = [n] = \{1, \dots, n\}$ we use abbreviations $\mathcal{A}_n$ and $\mathfrak{A}_n$. A* property *of CI structures is an element $\mathfrak{p}$ of the* property lattice

$$\mathfrak{P} \coloneqq \bigtimes_{n=1}^{\infty} 2^{\mathfrak{A}_n}.$$

A property $\mathfrak{p}$ consists of one set $\mathfrak{p}(n) \subseteq \mathfrak{A}_n$ per finite cardinality n. This is the set of CI structures over $[n]$ which "have property $\mathfrak{p}$". CI structures $\mathcal{L}$ and $\mathcal{M}$ over N and M, respectively, are *isomorphic* if there is a bijection $\pi : N \to M$ such that under the induced map $\mathcal{M} = \pi(\mathcal{L})$. We are only interested in properties which are invariant under isomorphy. Hence, the choice of ground sets $[n]$ presents no restriction. Moreover, we freely identify isomorphic CI structures in the following. In particular, each k-element subset $K \subseteq [n]$ will be tacitly identified with $[k]$ and we use notation such as $\mathfrak{p}(K)$.

Example 2. *By the localization rule* (L′)*, the well-known semigraphoid axioms of Pearl and Paz (1985) reduce to the single inference rule*

$$[i \perp\!\!\!\perp j \mid L] \wedge [i \perp\!\!\!\perp k \mid jL] \Rightarrow [i \perp\!\!\!\perp j \mid kL] \wedge [i \perp\!\!\!\perp k \mid L]. \tag{S}$$

Being a semigraphoid is a property defined by

$$\mathfrak{sg}(n) \coloneqq \{ \mathcal{L} \subseteq \mathcal{A}_n : \text{(S) holds for } \mathcal{L} \text{ for all distinct } i, j, k \in [n] \text{ and } L \subseteq [n] \setminus ijk \}.$$

Being realizable by a Gaussian distribution is another property

$$\mathfrak{g}^+(n) \coloneqq \{ [\![\Sigma]\!] \in \mathcal{A}_n : \Sigma \in \mathrm{PD}_n \}.$$

Both are closed under restriction, which can be expressed as follows: for every $\mathcal{L} \in \mathfrak{p}(N)$ and every $K \subseteq N$ we have $\mathcal{L}|_K \coloneqq \mathcal{L} \cap \mathcal{A}_K \in \mathfrak{p}(K)$.

The property lattice is equipped with a natural order relation of component-wise set inclusion from the boolean lattices $2^{\mathfrak{A}_n}$. This order relation $\le$ compares properties by generality: if $\mathfrak{p} \le \mathfrak{q}$, then for all $n \ge 1$ we have $\mathfrak{p}(n) \subseteq \mathfrak{q}(n)$, and $\mathfrak{p}$ is *sufficient* for $\mathfrak{q}$ and, equivalently, $\mathfrak{q}$ is *necessary* for $\mathfrak{p}$.

Definition 2. *Let $\mathfrak{p}$ be a property of CI structures. The* selfadhesion *$\mathfrak{p}^{\mathsf{sa}}(N)$ of $\mathfrak{p}$ is the set of CI structures $\mathcal{L}$ such that for every $L \subseteq N$ together with an L-copy M of N there exists $\overline{\mathcal{L}} \in \mathfrak{p}(NM)$ satisfying the two conditions:*

— $\overline{\mathcal{L}}|_N = \mathcal{L} = \overline{\mathcal{L}}|_M$*, and*
— $[N \perp\!\!\!\perp M \mid L] \in \overline{\mathcal{L}}$.

A property is selfadhesive *if $\mathfrak{p} = \mathfrak{p}^{\mathsf{sa}}$.*

The following is a direct consequence of Theorem 1:

Lemma 1. *The property $\mathfrak{g}^+$ of being regular Gaussian is selfadhesive.* □

Lemma 2. *The operator* $\cdot^{\mathsf{sa}}$ *is recessive and monotone on the property lattice.*

Proof. Let $\mathfrak{p}$ be a property and $\mathcal{L} \in \mathfrak{p}^{\mathsf{sa}}(N)$. In particular, $\mathcal{L}$ is selfadhesive with respect to $\mathfrak{p}$ at $L = N$. The L-copy M of N in the definition must be $M = N$ and it follows that $\mathcal{L} \in \mathfrak{p}(NM) = \mathfrak{p}(N)$. This proves recessiveness $\mathfrak{p}^{\mathsf{sa}} \le \mathfrak{p}$. For monotonicity, let $\mathfrak{p} \le \mathfrak{q}$ and $\mathcal{L}$ in $\mathfrak{p}^{\mathsf{sa}}(N)$. Then for every L with L-copy M of N there exist a certificate for the existence of $\mathcal{L}$ in $\mathfrak{p}^{\mathsf{sa}}$. This certificate lives in $\mathfrak{p}(NM) \subseteq \mathfrak{q}(NM)$ which proves $\mathcal{L} \in \mathfrak{q}^{\mathsf{sa}}(N)$. □

Thus, from monotonicity and the fact that $\mathfrak{g}^+$ is a fixed point of selfadhesion, we obtain the following crucial lemma which states that a property which is necessary for Gaussianity remains necessary after selfadhesion. Since selfadhesion makes properties more specific, this allows one to take known necessary properties of Gaussian CI and to derive new, stronger properties from them.

Lemma 3. *If* $\mathfrak{g}^+ \le \mathfrak{p}$*, then* $\mathfrak{g}^+ \le \mathfrak{p}^{\mathsf{sa}}$*.* □

Whether or not a CI structure $\mathcal{L} \subseteq \mathcal{A}_N$ is in $\mathfrak{p}^{\mathsf{sa}}$ can be checked if an oracle $\mathsf{p}(\widetilde{\mathcal{L}})$ for the property $\mathfrak{p}$ is available. This oracle receives a *partially defined* CI structure $\widetilde{\mathcal{L}}$ over N, i.e., a set of CI statements or negated CI statements specifying constraints on some statements from $\mathcal{A}_N$. Then the oracle p decides if $\widetilde{\mathcal{L}}$ can be extended to a member of $\mathfrak{p}(N)$.

Algorithm 1 Blackbox selfadhesion membership test

function in-selfadhesion$(\mathcal{L}, \mathsf{p})$ ▷ tests if $\mathcal{L} \in \mathfrak{p}^{\mathsf{sa}}(N)$
  for all $L \subseteq N$ **do**
    $(M, \pi) \leftarrow L$-copy of N with bijection $\pi : N \to M$
    $\widetilde{\mathcal{L}} \leftarrow \emptyset$
    for all $s \in \mathcal{A}_N$ **do**
      $\widetilde{\mathcal{L}} \leftarrow \widetilde{\mathcal{L}} \cup \{\ s,\ \pi(s)\ \}$ **if** $s \in \mathcal{L}$ **or else**
      $\widetilde{\mathcal{L}} \leftarrow \widetilde{\mathcal{L}} \cup \{\neg s, \neg\pi(s)\}$ **if** $s \notin \mathcal{L}$
    end for
    $\widetilde{\mathcal{L}} \leftarrow \widetilde{\mathcal{L}} \cup \{[N \perp\!\!\!\perp M \mid L]\}$ ▷ or equivalent statements via (L) or (L′)
    return `false` **if** $\mathsf{p}(\widetilde{\mathcal{L}}) =$ `false`
  end for
  return `true`
end function

Remark 2. *The proof of Lemma 2 shows that a CI structure* $\mathcal{L}$ *satisfies selfadhesivity with respect to* $\mathfrak{p}$ *at* $L = N$ *if and only if* $\mathcal{L}$ *has property* $\mathfrak{p}$*. In the other extreme case, every structure in* $\mathfrak{p}$ *is selfadhesive at* $L = \emptyset$ *if* $\mathfrak{p}$ *is closed under direct sums. Many useful properties are closed under direct sums because this operation mimics the independent joining of two random vectors; see Matúš (2004).*

Before proceeding with Algorithm 1 to two practically tractable necessary conditions, we record the following important question concerning the selfadhesion operator:

Question 1. *Does* $\cdot^{\mathsf{sa}}$ *stabilize after the first application to "well-behaved" properties like* $\mathfrak{sg}$*? Under which assumptions on a property does it stabilize eventually?*

4.1 Structural semigraphoids

It is easy to see that every Gaussian CI structure $\mathcal{L} = [\![\Sigma]\!]$ can also be obtained from the *correlation matrix* Σ' of the original distribution Σ. Hence, we may assume that Σ is a correlation matrix. In that case, the *multiinformation vector* of Σ is the map $m_\Sigma : 2^N \to \mathbb{R}$ given by $m_\Sigma(K) \coloneqq -1/2 \log\det \Sigma_K$. This function satisfies $m_\Sigma(\emptyset) = m_\Sigma(i) = 0$ for all $i \in N$ and it is supermodular by the Koteljanskii inequality; see Johnson and Barrett (1993). Similarly to entropy vectors, the equality condition in these inequalities characterizes conditional independence: $\triangle(ij|K) \cdot m_\Sigma = 0 \;\Leftrightarrow\; [i \perp\!\!\!\perp j \mid K] \in [\![\Sigma]\!]$.

In the nomenclature of Studený (2005), Chapter 5, m_Σ is an *ℓ-standardized supermodular function.* The functions having these two properties form a rational, polyhedral cone $\mathbf{S}_N$ of codimension $|N| + 1$ in $\mathbb{R}^{2^N}$. Its facets are precisely given by the supermodular inequalities $\triangle(ij|K) \le 0$ for all elementary CI statements $[i \perp\!\!\!\perp j \mid K] \in \mathcal{A}_N$. Since the facets of this cone are in bijection with CI statements, it is natural to identify faces (intersections of facets) dually with CI structures (unions of CI statements). The property of CI structures defined by arising from a face of $\mathbf{S}_N$ is that of *structural semigraphoids*, denoted by $\mathfrak{sg}_*$, and it is necessary for $\mathfrak{g}^+$ since every Gaussian CI structure $[\![\Sigma]\!]$ is associated with the unique face on which $m_\Sigma \in \mathbf{S}_N$ lies in the relative interior.

Deciding whether a partially defined CI structure $\widetilde{\mathcal{L}}$ is consistent with this property is a question about the incidence structure of the face lattice of $\mathbf{S}_N$. Such questions reduce to the feasibility of a linear program as previously demonstrated by Bouckaert et al. (2010):

Algorithm 2 Structural semigraphoid consistency test

function is-structural($\widetilde{\mathcal{L}}$) ▷ tests if $\widetilde{\mathcal{L}}$ can be extended to a member of $\mathfrak{sg}_*(N)$
    $P \leftarrow \{\, m(\emptyset) = m(i) = 0$ for all $i \in N \,\}$
    for all $s \in \mathcal{A}_N$ **do**
        $P \leftarrow P \cup \{\, -\triangle(s) \cdot m = 0 \,\}$ **if** $s \in \widetilde{\mathcal{L}}$ **or else**
        $P \leftarrow P \cup \{\, -\triangle(s) \cdot m \ge 1 \,\}$ **if** $\neg s \in \widetilde{\mathcal{L}}$ **or else**
        $P \leftarrow P \cup \{\, -\triangle(s) \cdot m \ge 0 \,\}$
    end for
    return is-feasible(P) ▷ call an LP solver
end function

Equipped with this oracle for $\mathfrak{sg}_*$ Algorithm 1 can be applied to compute membership in $\mathfrak{sg}_*^{\mathsf{sa}}$. We use the *gaussoids* of Lněnička and Matúš (2007) as input because they are easily computable candidates for Gaussian CI structures; see also Boege et al. (2019). For $n = 4$ random variables, the gaussoids which are structural semigraphoids already coincide with the realizable Gaussian structures and selfadhesivity offers no improvement. This is no longer the case on five random variables:

Computation 1. *There are* 508 817 *gaussoids on* $n = 5$ *random variables modulo isomorphy. Of these* 336 838 *are structural semigraphoids and* 335 047 *of them are selfadhesive with respect to* $\mathfrak{sg}_*$.

4.2 Orientable gaussoids

Recall from Boege et al. (2019) that a gaussoid is *orientable* if it is the support of an oriented gaussoid. Oriented gaussoids are a variant of CI structures in which every statement $[i \perp\!\!\!\perp j \mid K]$ has a sign $\{\, \mathtt{0}, \mathtt{+}, \mathtt{-} \,\}$ attached, indicating conditional independence, positive or negative partial correlation, respectively. Oriented gaussoids are axiomatically defined and therefore SAT solvers are ideally suited to decide the consistency of a partially defined CI structure with these axioms. The property of orientability, denoted $\mathfrak{o}$, is obtained from the set of oriented gaussoids by mapping all CI statements oriented as 0 to elements of a CI structure and all statements oriented + or - to non-elements. To facilitate orientability testing, one allocates two boolean variables $V_s^{\mathtt{0}}$ and $V_s^{\mathtt{+}}$ for every CI statement s. The former indicates whether s is 0 or not while the latter indicates, provided that $V_s^{\mathtt{0}}$ is false, if s is + or -. Further details about oriented gaussoids, their axioms and use of SAT solvers for CI inference are available in Boege et al. (2019).

Algorithm 3 Orientable gaussoid consistency test

function is-orientable($\widetilde{\mathcal{L}}$) ▷ tests if $\widetilde{\mathcal{L}}$ can be extended to a member of $\mathfrak{o}(N)$
  $\varphi \leftarrow$ oriented-gaussoid-axioms(N)
  for all $s \in \mathcal{A}_N$ **do**
    $\varphi \leftarrow \varphi \wedge [V_s^{\mathtt{0}} = \mathtt{true}]$ **if** $s \in \widetilde{\mathcal{L}}$ **or else**
    $\varphi \leftarrow \varphi \wedge [V_s^{\mathtt{0}} = \mathtt{false}]$ **if** $\neg s \in \widetilde{\mathcal{L}}$
    $\varphi \leftarrow \varphi \wedge [V_s^{\mathtt{0}} = \mathtt{true} \Rightarrow V_s^{\mathtt{+}} = \mathtt{false}]$ ▷ there are only three signs $\{\, \mathtt{0}, \mathtt{+}, \mathtt{-} \,\}$
  end for
  return is-satisfiable(φ) ▷ call a SAT solver
end function

Computation 2. *All orientable gaussoids on $n = 4$ are Gaussian. Of the* 508 817 *isomorphy classes of gaussoids on $n = 5$ precisely* 175 215 *are orientable and* 168 010 *are selfadhesive with respect to orientability.*

4.3 Structural orientable gaussoids

The meet property $\mathfrak{sg}_* \wedge \mathfrak{o}$ of structural semigraphoids and orientable gaussoids is likewise necessary for Gaussianity and an oracle for it can be combined from the oracles of its two constituents. Its selfadhesion yields no improvement over apparently weaker properties:

Computation 3. *The properties $\mathfrak{sg}_* \wedge \mathfrak{o}$ and $\mathfrak{sg}_*^{\mathsf{sa}} \wedge \mathfrak{o}$ coincide at $n = 5$ with* 175 139 *isomorphy types. On the other hand, $\mathfrak{sg}_* \wedge \mathfrak{o}^{\mathsf{sa}}$, $\mathfrak{sg}_*^{\mathsf{sa}} \wedge \mathfrak{o}^{\mathsf{sa}}$ and $(\mathfrak{sg}_* \wedge \mathfrak{o})^{\mathsf{sa}}$ coincide at $n = 5$ with* 167 989 *types.*

Up to a few isolated examples in the literature, this represents the currently best known upper bound in the classification of realizable Gaussian conditional independence structures on five random variables. Examination of the difference $(\mathfrak{sg}_* \wedge \mathfrak{o})(5) \setminus (\mathfrak{sg}_* \wedge \mathfrak{o})^{\mathsf{sa}}(5)$ reveals new axioms for Gaussian CI beyond structural semigraphoids and orientability, e.g.:

$$[i \perp\!\!\!\perp j \mid km] \wedge [i \perp\!\!\!\perp m \mid l] \wedge [j \perp\!\!\!\perp k \mid i] \wedge [j \perp\!\!\!\perp m] \wedge [k \perp\!\!\!\perp l] \;\Rightarrow\; [i \perp\!\!\!\perp j].$$

Mathematical software and data repository. `SoPlex v4.0.0` was used to solve rational linear programs exactly; see Gleixner et al. (2012, 2015, 2018). To check orientability, we used the incremental `SAT` solver `CaDiCaL v1.3.1` by Biere (2019) and to enumerate satisfying assignments the `AllSAT` solver `nbc_minisat_all v1.0.2` by Toda and Soh (2016). The source code and results for all computations are available on the supplementary MathRepo website of the MPI-MiS:

https://mathrepo.mis.mpg.de/SelfadhesiveGaussianCI/.

Acknowledgement. I would like to thank the Otto-von-Guericke-Universität Magdeburg and the Max Planck Institute for Mathematics in the Sciences for providing me with the resources to carry out the computations whose results are presented here.

References

A. Biere. CaDiCaL at the SAT Race 2019. In M. Heule, M. Järvisalo, and M. Suda, editors, *Proc. of SAT Race 2019 – Solver and Benchmark Descriptions*, volume B-2019-1 of *Department of Computer Science Series of Publications B*, pages 8–9. University of Helsinki, 2019.

T. Boege. Gaussoids are two-antecedental approximations of Gaussian conditional independence structures. *Ann. Math. Artif. Intell.*, 2021. doi: 10.1007/s10472-021-09780-0.

T. Boege, A. D'Alì, T. Kahle, and B. Sturmfels. The Geometry of Gaussoids. *Found. Comput. Math.*, 19(4):775–812, 2019. doi: 10.1007/s10208-018-9396-x.

R. Bouckaert, R. Hemmecke, S. Lindner, and M. Studený. Efficient algorithms for conditional independence inference. *J. Mach. Learn. Res.*, 11:3453–3479, 2010.

T. H. Chan. Balanced information inequalities. *IEEE Trans. Inf. Theory*, 49(12):3261–3267, 2003. doi: 10.1109/TIT.2003.820037.

L. Csirmaz. Book inequalities. *IEEE Trans. Inf. Theory*, 60(11):6811–6818, 2014. doi: 10.1109/TIT.2014.2352273.

A. Gleixner, D. Steffy, and K. Wolter. Improving the accuracy of linear programming solvers with iterative refinement. Technical Report 12-19, ZIB, Takustr. 7, 14195 Berlin, 2012.

A. Gleixner, D. Steffy, and K. Wolter. Iterative refinement for linear programming. Technical Report 15-15, ZIB, Takustr. 7, 14195 Berlin, 2015.

A. Gleixner, M. Bastubbe, L. Eifler, T. Gally, G. Gamrath, R. L. Gottwald, G. Hendel, C. Hojny, T. Koch, M. Lübbecke, S. J. Maher, M. Miltenberger, B. Müller, M. Pfetsch, C. Puchert, D. Rehfeldt, F. Schlösser, C. Schubert, F. Serrano, Y. Shinano, J. M. Viernickel, M. Walter, F. Wegscheider, J. T. Witt, and J. Witzig. The SCIP Optimization Suite 6.0. Technical Report 18-26, ZIB, Takustr. 7, 14195 Berlin, 2018.

C. R. Johnson and W. W. Barrett. Determinantal inequalities for positive definite matrices. *Discrete Math.*, 119(1-3):97–106, 1993. doi: 10.1016/0012-365X(93)90119-E.

R. Lněnička. On the tightness of the Zhang-Yeung inequality for Gaussian vectors. *Commun. Inf. Syst.*, 3(1):41–46, 2003. doi: 10.4310/CIS.2003.v3.n1.a3.

R. Lněnička and F. Matúš. On Gaussian conditional independence structures. *Kybernetika*, 43(3):327–342, 2007.

F. Matúš. Ascending and descending conditional independence relations. In *Transactions of the 11th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, volume B, pages 189–200, 1992.

F. Matúš. Probabilistic conditional independence structures and matroid theory: background. *Int. J. Gen. Syst.*, 22:185–196, 1994. doi: 10.1080/03081079308935205.

F. Matúš. Towards classification of semigraphoids. *Discrete Math.*, 277(1):115–145, 2004. doi: 10.1016/S0012-365X(03)00155-9.

F. Matúš. Adhesivity of polymatroids. *Discrete Math.*, 307(21):2464–2477, 2007. doi: 10.1016/j.disc.2006.11.013.

J. Pearl and A. Paz. GRAPHOIDS: A graph-based logic for reasoning about relevance relations, or When would x tell you more about y if you already know z. Technical Report CSD-850038, UCLA Computer Science Department, 1985.

M. Studený. Structural semigraphoids. *Int. J. Gen. Syst.*, 22(2):207–217, 1994. doi: 10.1080/03081079308935207.

M. Studený. *Probabilistic Conditional Independence Structures.* Information Science and Statistics. Springer, 2005.

S. Sullivant. *Algebraic Statistics*, volume 194 of *Graduate Studies in Mathematics.* American Mathematical Society (AMS), 2018.

T. Toda and T. Soh. Implementing efficient all solutions SAT solvers. *J. Exp. Algorithm.*, 21:1.12:1–1.12:44, 2016. doi: 10.1145/2975585.

F. Zhang, editor. *The Schur complement and its applications.* Springer, 2005.

Z. Zhang and R. W. Yeung. On characterization of entropy function via information inequalities. *IEEE Trans. Inf. Theory*, 44(4):1440–1452, 1998. doi: 10.1109/18.681320.

Two Generalizations of the Semi-graphoid Rule of Probabilistic Independence

Janneke H. Bolt

Department of Mathematics and Computer Science,
Eindhoven University of Technology
j.h.bolt@tue.nl

Abstract

Probabilistic independence is a key concept in probability theory and statistics. For probabilistic independence a set of well known qualitative rules exists, the so-called semi-graphoid rules, which can, besides the rule of symmetry, be summarized into a single semi-graphoid rule. The rule system is not complete and an additional five rules were formulated. One of those rules was even generalized, which proofs that no finite rule system exists. In recent work, all five additional rules were (further) generalized. In this paper two generalizations of the semi-graphoid rule are given and two new rules of probabilistic independence are stated. The paper thereby contributes to the insights into the structural properties of probabilistic independence, and to an enhanced description of probabilistic independence.

Keywords: probabilistic independence; rules of probabilistic independence; semi-graphoid rules

1 Introduction

Probabilistic independence is a key concept in probability theory and statistics and plays an important role in probabilistic reasoning and probabilistic graphical models. In (Pearl, 1988), a set of sound qualitative rules of probabilistic independence is given; the so-called semi-graphoid rules. Besides the rule that captures the symmetry property of probabilistic independence, these rules can be summarized into a single semi-graphoid rule (Studený, 1989a). Pearl conjectured that the semi-graphoid rules would be complete for probabilistic independence, however, in (Studený, 1989b) a new rule of probabilistic independence is formulated. The semi-graphoid rules thus are incomplete and a set of independence statements that is closed under these rules, may lack statements that are enforced given this set. In (Studený, 1994), another four new rules are stated and in (Studený, 1992; Studený and Vejnarová, 1998) the authors show that there is no finite complete system of rules of probabilistic independence by generalizing one of the new rules to an unlimited number of variables. For proving the correctness of the rules properties of the conditional mutual information between (sets of) variables and the relation between conditional mutual information and the multi-information function are used.

In recent work, Bolt and van der Gaag (2021) show that not just one, but that all five new rules can be generalized to an unlimited number of variables. These generalized rules give enhanced insights into the structural properties of probabilistic independence and give an enhanced description of probabilistic independence. In this paper two generalization of the semi-graphoid rule are formulated and two additional rules of probabilistic independence are given. The paper thereby further contributes to the insights in, and the description of probabilistic independence.

2 Preliminaries

2.1 Probabilistic and semi-graphoid independence relations

Throughout the paper, a set V of discrete random variables with subsets A, B, C, ... is considered. Set union is noted by concatenation of the sets, so $A \cup B$ is written as AB. A triplet $\langle A, B \mid C \rangle$, with A, B, C pairwise disjoint subsets of V and A and B nonempty, states that the sets of variables A and B are probabilistic independent given observations for the variables in set C. An elementary triplet is a triplet with A and B singletons. Any set of triplets over V is called an independence relation. A probabilistic independence relation is an independence relation for which a matching probability distribution exists.

Pearl (1988) states four rules of probabilistic independence, which sum up in the following two rules (Studený, 1989a):

$A1$: $\langle A, B \mid C \rangle \;\leftrightarrow\; \langle B, A \mid C \rangle$

$A2$: $\langle AB, C \mid D \rangle \;\leftrightarrow\; \langle A, C \mid BD \rangle \wedge \langle B, C \mid D \rangle$

Any independence relation closed under these two rules is called a semi-graphoid independence relation.

Matúš (1992) argues that any semi-graphoid independence relation is fully captured by its elementary triplets. He moreover considers the first two positions of a triplet as unordered and alternatively defines a semi-graphoid independence relation as a set of elementary triplets that is closed under the rule:

$A2'$: $\langle A, C \mid D \rangle \wedge \langle B, C \mid AD \rangle \;\leftrightarrow\; \langle A, C \mid BD \rangle \wedge \langle B, C \mid D \rangle$

The semi-graphoid rules are sound with respect to probabilistic independence relations. The system of rules, however, is not complete, as is shown by Studený (1989b) by stating a new rule of probabilistic independence. The correctness of this new rule is proved by using the relation between condition mutual information and the so-called multiinformation function. This is further discussed in the next section.

In (Studený, 1994) another four new rules are stated and in (Studený, 1992; Studený and Vejnarová, 1998) the authors show that there is no finite complete system of rule of probabilistic independence by a generalizing of one of the new rules to an unlimited number of variables. In (Bolt and van der Gaag, 2021) it is shown that all rules given in (Studený, 1994) can be generalized to an unlimited number of variables.

2.2 Mutual information, multiinformation and probabilistic independence

In proving the correctness of the new rules of probabilistic independence, the relation between the mutual conditional information and the so-called multiinformation function is used, as discussed below (Studený, 1989b; Studený and Vejnarová, 1998).

Given a probability distribution Pr over V, the mutual information of two sets of random variables A and B in the context of a third set C, noted $I(A;B|C)$, is a measure of the mutual dependence between A and B in the context of C (see for example (Yeung, 2002)). The conditional mutual information has as properties that for any A, B, C

- $I(A;B|C) \geq 0$;
- $I(A;B|C) = 0$ iff $\langle A, B \mid C \rangle$.

The multiinformation function induced by a probability distribution over V is a real function $M\colon 2^V \to [0, \infty)$ on the power set of V. The mutual conditional information is related to the multiinformation function by:

- $I(A;B \mid C) = M(ABC) + M(C) - M(AC) - M(BC)$

We thus have that:

- $M(ABC) + M(C) - M(AC) - M(BC) \geq 0$;
- $M(ABC) + M(C) - M(AC) - M(BC) = 0$ iff $\langle A, B \mid C \rangle$.

This relation enables straightforward proofs for rules of independence: a rule is sound if the multiinformation terms of its set of premise triplets can be converted into the multiinformation terms of its set of consequent triplets. Below as example a proof of semi-graphoid rule $A2$.

Example 1 *The probabilistic soundness of* $\langle AB, C \mid D \rangle \leftrightarrow \langle A, C \mid BD \rangle \wedge \langle B, C \mid D \rangle$ *is proved as follows:*

We have that $\langle AB, C \mid D \rangle$ *is a valid independence statement if and only if*

$$\begin{aligned} 0 &= M(ABCD) + M(D) - M(ABD) - M(CD) \ \Leftrightarrow \\ 0 &= M(ABCD) + M(D) - M(ABD) - M(CD) \\ &\quad + M(BD) - M(BD) + M(BCD) - M(BCD) \ \Leftrightarrow \\ 0 &= M(ABCD) + M(BD) - M(ABD) - M(BCD) \\ &\quad + M(BCD) + M(D) - M(BD) - M(CD) \ \Leftrightarrow \end{aligned}$$

which is true if and only if $\langle A, C \mid BD \rangle$ *and* $\langle B, C \mid D \rangle$ *are valid independence statements.*

The last step is based on the fact that the conditional mutual information for any three sets of variables is larger than or equal to 0.

3 Two generalizations of rule $A2'$

In this section in Proposition 1 and 2, two different generalizations of rule $A2'$ are given. Proofs of the propositions are provided in the appendix.

First note that the triplets of rule $A1$, the triplets of rule $A2$ and the triplets of rule $A2'$ share a set of conditioning variables. (The set C in rule $A1$ and the set D in rules $A2$ and $A2'$.) In the generalizations given in this paragraph and in the two new rules in the next paragraph such a shared set is omitted for clarity of exposition. A shared condition set can be added to the triplets of a rule without affecting its validity. The proof of a rule's validity with or without such a set is fully analogous.

Both generalizations stated in this section involve the sets A and B and a set $\mathbf{C}$ of sets C_i, with i an odd number. For each C_i two triplets are found both in the premise and in the consequent of the rules. One triplet with A as first argument and C_i as second argument and one triplet with B as first argument and C_i as second argument. The sets $\mathbf{C} \setminus C_i$ are distributed over the third arguments of those two triplets; in one of those triplets supplemented with the set A or B. The two rules differ in the specific composition of the third arguments of their triplets

Proposition 1 *Let $A, B, C_1, \ldots, C_n$ with $n \geq 1$, n is odd, be non-empty, mutually disjoint sets of variables. Then (taking $C_i \cdots C_{i-1} := \varnothing$),*

G2a:

$$\bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ odd}}} \big[\langle A, C_i \mid C_1 \cdots C_{i-1}\rangle \wedge \langle B, C_i \mid A\, C_{i+1} \cdots C_n\rangle\big] \wedge \bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ even}}} \big[\langle A, C_i \mid B\, C_1 \cdots C_{i-1}\rangle \wedge \langle B, C_i \mid C_{i+1} \cdots C_n\rangle\big] \leftrightarrow$$

$$\bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ odd}}} \big[\langle A, C_i \mid B\, C_{i+1} \cdots C_n\rangle \wedge \langle B, C_i \mid C_1 \cdots C_{i-1}\rangle\big] \wedge \bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ even}}} \big[\langle A, C_i \mid C_{i+1} \cdots C_n\rangle \wedge \langle B, C_i \mid A\, C_1 \cdots C_{i-1}\rangle\big]$$

is a sound rule of probabilistic independence.

For $n = 1$ (with an additional conditioning set D) the rule equals rule $A2'$. For $n = 3$ the rule states that

$$\langle A, C_1 \mid \varnothing\rangle \wedge \langle A, C_3 \mid C_1 C_2\rangle \wedge \langle B, C_1 \mid A\, C_2 C_3\rangle \wedge \langle B, C_3 \mid A\rangle \wedge \langle A, C_2 \mid B\, C_1\rangle \wedge \langle B, C_2 \mid C_3\rangle \leftrightarrow$$

$$\langle A, C_1 \mid B\, C_2 C_3\rangle \wedge \langle A, C_3 \mid B\rangle \wedge \langle B, C_1 \mid \varnothing\rangle \wedge \langle B, C_3 \mid C_1 C_2\rangle \wedge \langle A, C_2 \mid C_3\rangle \wedge \langle B, C_2 \mid A\, C_1\rangle$$

In Figure 1 the structure of $G2$a is clarified for $n = 5$ by a more synoptic representation. Each $X \overset{Z}{\text{---}} Y$ in this figure represents a triplet $\langle X, Y \mid Z\rangle$.

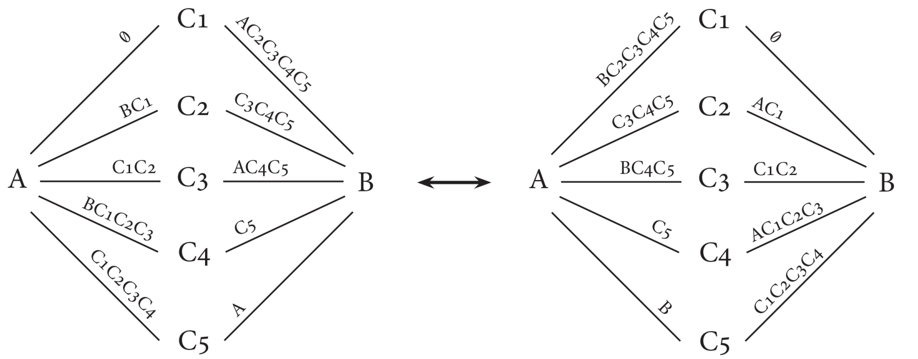


Figure 1: The structure of rule $G2$a for $n = 5$.

Proposition 2 *Let $A, B, C_1, \ldots, C_n$ with $n \geq 1$, n is odd, be non-empty, mutually disjoint sets of variables. Then (taking $C_i \cdots C_{i-1} := \varnothing$)*

G2b:

$$\bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ odd}}} \left[\langle A, C_i \mid C_1 \cdots C_{i-1} \rangle \;\wedge\; \langle B, C_i \mid A\, C_{i+1} \cdots C_n \rangle \right] \;\wedge \bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ even}}} \left[\langle A, C_i \mid B\, C_{i+1} \cdots C_n \rangle \;\wedge\; \langle B, C_i \mid C_1 \cdots C_{i-1} \rangle \right] \;\leftrightarrow$$

$$\bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ odd}}} \left[\langle A, C_i \mid B\, C_1 \cdots C_{i-1} \rangle \;\wedge\; \langle B, C_i \mid C_{i+1} \cdots C_n \rangle \right] \;\wedge \bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ even}}} \left[\langle A, C_i \mid C_{i+1} \cdots C_n \rangle \;\wedge\; \langle B, C_i \mid A\, C_1 \cdots C_{i-1} \rangle \right]$$

is a sound rule of probabilistic independence.

For $n = 1$ (with an additional conditioning set D) the rule equals rule $A2'$. For $n = 3$ the rule states that

$$\langle A, C_1 \mid \varnothing \rangle \;\wedge\; \langle A, C_3 \mid C_1 C_2 \rangle \;\wedge\; \langle B, C_1 \mid A\, C_2 C_3 \rangle \;\wedge\; \langle B, C_3 \mid A \rangle \;\wedge$$
$$\langle A, C_2 \mid B\, C_3 \rangle \;\wedge\; \langle B, C_2 \mid C_1 \rangle \quad \leftrightarrow$$
$$\langle A, C_1 \mid B \rangle \;\wedge\; \langle A, C_3 \mid BC_1 C_2 \rangle \;\wedge\; \langle B, C_1 \mid C_2 C_3 \rangle \;\wedge\; \langle B, C_3 \mid \varnothing \rangle \;\wedge$$
$$\langle A, C_2 \mid C_3 \rangle \;\wedge\; \langle B, C_2 \mid A\, C_1 \rangle$$

In Figure 2 the structure of rule $G2$b is clarified for $n = 5$.

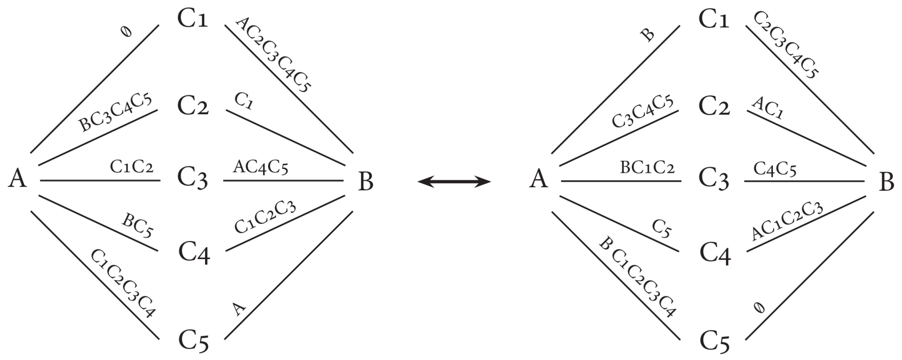


Figure 2: The structure of rule $G2$b for $n = 5$.

4 Two additional rules of probabilistic independence

Rules $A1$ and $A2$ plus the generalized rules given in Bolt and van der Gaag (2021) and in this paper are not exhaustive for the rules that can be found using the method described in Section 2.2. Below two more rules of probabilistic independence are given.

Proposition 3 *Let A, B, C, D, E, be non-empty, mutually disjoint sets of variables. Then,*

$$A8: \quad \langle A, C \mid \varnothing \rangle \wedge \langle A, D \mid CE \rangle \wedge \langle A, E \mid BC \rangle \wedge \langle B, C \mid ADE \rangle \wedge \langle B, D \mid E \rangle \wedge \langle B, E \mid A \rangle \quad \leftrightarrow$$

$$\langle A, C \mid BDE \rangle \wedge \langle A, D \mid E \rangle \wedge \langle A, E \mid B \rangle \wedge \langle B, C \mid \varnothing \rangle \wedge \langle B, D \mid CE \rangle \wedge \langle B, E \mid AC \rangle$$

is a sound rule of probabilistic independence.

Proof. The proposition can be proved straightforwardly by using the method described in Section 2.2. □

The structure of the rule is clarified in Figure 3.

Proposition 4 *Let A, B, C, D, E, be non-empty, mutually disjoint sets of variables. Then,*

$$A9: \quad \langle A, C \mid \varnothing \rangle \wedge \langle A, D \mid CE \rangle \wedge \langle A, E \mid BC \rangle \wedge \langle B, C \mid D \rangle \wedge \langle B, D \mid A \rangle \wedge \langle B, E \mid ACD \rangle \quad \leftrightarrow$$

$$\langle A, C \mid D \rangle \wedge \langle A, D \mid B \rangle \wedge \langle A, E \mid BCD \rangle \wedge \langle B, C \mid \emptyset \rangle \wedge \langle B, D \mid CE \rangle \wedge \langle B, E \mid AC \rangle$$

is a sound rule of probabilistic independence.

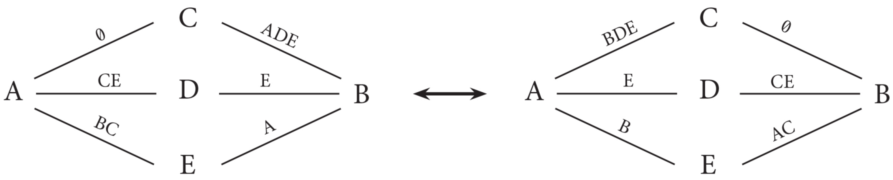


Figure 3: The structure of rule $A8$.

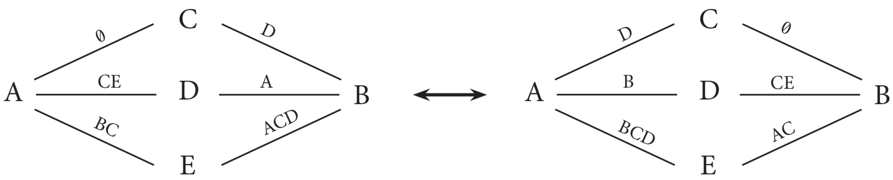


Figure 4: The structure of rule $A9$.

Proof. The proposition can be proved straightforwardly by using the method described in Section 2.2. □

The structure of the rule is clarified in Figure 4.

5 Conclusions and future research

In this paper two generalization for the semi-graphoid rule of probabilistic independence are stated and two new rules of probabilistic independence are given. The paper thereby contributes to the insights in probabilistic independence and to the description of probabilistic independence by a qualitative rule system. The paper also shows that more than one generalization of a single rule may exist. The correctness of the new rules is proved by a method based on the relation between conditional mutual independence and the concept of multiinformation. An obvious question for future research is whether rules $A8$ and $A9$, can be generalized as well. Another, more fundamental, question is whether the number of (generalized) rules that can be found by the proof method that is used is unlimited.

References

J. H. Bolt and L. C. van der Gaag. Generalized rules of probabilistic independence. In J. Vejnarová and N. Wilson, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 590–602, Cham, 2021. Springer International Publishing. ISBN 978-3-030-86772-0.

F. Matúš. Ascending and descending conditional independence relations. In S. Kubik and J.A. Visek, editors, *Proceedings of the Joint Session of the 11th Prague Confer-*

ence on Asymptotic Statistics and the 13th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, pages 189–200. Kluwer, 1992.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, 1988.

M. Studený. Attempts at axiomatic description of conditional independence. *Kybernetika*, 25:72–79, 1989a.

M. Studený. Multiinformation and the problem of characterization of conditional independence relations. *Problems of Control and Information Theory*, 18, 01 1989b.

M. Studený. Conditional independence relations have no finite complete characterization. In *Information Theory, Statistical Decision Functions and Random Processes. Transactions of the 11th Prague Conference vol. B*, pages 377–396. Kluwer, Dordrecht - Boston - London (also Academia, Prague), 1992.

M. Studený. Structural semigraphoids. *Int. J. General Systems*, 22:207–217, 1994.

M. Studený and J. Vejnarová. The multiinformation function as a tool for measuring stochastic dependence. In *Learning in Graphical Models*, pages 261–298. Kluwer Academic Publishers, 1998.

R. Yeung. *A First Course in Information Theory.* Kluwer Academic Publishers, Boston/Dordrecht/London, 2002.

Appendix: Proofs of Proposition 1 and Proposition 2

Both propositions are proved using the method described in Section 2.2. In both proofs $\{i, \ldots, i-2\} := \varnothing$ and

$$
\begin{aligned}
m_1 &= M(C_1 \cdots C_{i-1}) & m_9 &= M(BC_1 \cdots C_{i-1}) \\
m_2 &= M(C_1 \cdots C_i) & m_{10} &= M(BC_1 \cdots C_i) \\
m_3 &= M(C_i \cdots C_n) & m_{11} &= M(BC_i \cdots C_n) \\
m_4 &= M(C_{i+1} \cdots C_n) & m_{12} &= M(BC_{i+1} \cdots C_n) \\
m_5 &= M(AC_1 \cdots C_{i-1}) & m_{13} &= M(ABC_1 \cdots C_{i-1}) \\
m_6 &= M(AC_1 \cdots C_i) & m_{14} &= M(ABC_1 \cdots C_i) \\
m_7 &= M(AC_i \cdots C_n) & m_{15} &= M(ABC_i \cdots C_n) \\
m_8 &= M(AC_{i+1} \cdots C_n) & m_{16} &= M(ABC_{i+1} \cdots C_n)
\end{aligned}
$$

Proposition 1

We have that

$$\bigwedge_{\substack{i\,\in\,\{1,\ldots,n\},\\ i\text{ odd}}} \left[\,\langle A, C_i \mid C_1\cdots C_{i-1}\rangle \;\wedge\; \langle B, C_i \mid A\,C_{i+1}\cdots C_n\rangle\,\right] \;\wedge\; \bigwedge_{\substack{i\,\in\,\{1,\ldots,n\},\\ i\text{ even}}} \left[\,\langle A, C_i \mid B\,C_1\cdots C_{i-1}\rangle \;\wedge\; \langle B, C_i \mid C_{i+1}\cdots C_n\rangle\,\right]$$

are valid independence statements if and only if

$$\begin{aligned} 0 \;=\; & \sum_{i\,\in\,\{1,\ldots,n\},\, i\text{ odd}} \left[\,m_6+m_1-m_5-m_2+m_{15}+m_8-m_{16}-m_7\,\right] \\ & + \sum_{i\,\in\,\{1,\ldots,n\},\, i\text{ even}} \left[\,m_{14}+m_9-m_{13}-m_{10}+m_{11}+m_4-m_{12}-m_3\,\right] \quad \Leftrightarrow \end{aligned}$$

$$\begin{aligned} 0 \;=\; & \sum_{i\,\in\,\{1,\ldots,n\},\, i\text{ odd}} \left[\,m_1-m_2+m_{15}-m_{16}\,\right] \\ & + \sum_{i\,\in\,\{1,\ldots,n-2\},\, i\text{ odd}} \left[\,m_6+m_8\,\right] + M(AC_1\cdots C_n) + M(A) \\ & + \sum_{i\,\in\,\{3,\ldots,n\},\, i\text{ odd}} \left[\,-m_5-m_7\,\right] - M(A) - M(AC_1\cdots C_n) \\ & + \sum_{i\,\in\,\{1,\ldots,n\},\, i\text{ even}} \left[\,m_{14}+m_9-m_{13}-m_{10}+m_{11}+m_4-m_{12}-m_3\,\right] \end{aligned}$$

Since

$$\sum_{i\,\in\,\{1,\ldots,n-2\},\, i\text{ odd}} \left[\,m_6+m_8\,\right] \;=\; \sum_{i\,\in\,\{1,\ldots,n\},\, i\text{ even}} \left[\,m_5+m_7\,\right]$$

$$\sum_{i\,\in\,\{3,\ldots,n\},\, i\text{ odd}} \left[\,-m_5-m_7\,\right] \;=\; \sum_{i\,\in\,\{1,\ldots,n\},\, i\text{ even}} \left[\,-m_6-m_8\,\right]$$

$$\sum_{i\,\in\,\{1,\ldots,n\},\, i\text{ even}} \left[\,m_9+m_{11}\,\right] \;=\; \sum_{i\,\in\,\{1,\ldots,n-2\},\, i\text{ odd}} \left[\,m_{10}+m_{12}\,\right]$$

$$\sum_{i\,\in\,\{1,\ldots,n\},\, i\text{ even}} \left[\,-m_{10}-m_{12}\,\right] \;=\; \sum_{i\,\in\,\{3,\ldots,n\},\, i\text{ odd}} \left[\,-m_9-m_{11}\,\right]$$

and

$$M(AC_1\cdots C_n) + M(A) - M(A) - M(AC_1\cdots C_n) = 0 =$$
$$M(B) + M(BC_1\cdots C_n) - M(BC_1\cdots C_n) - M(B)$$

we find that

$$
\begin{aligned}
0 \;=\; & \sum_{i \in \{1,\ldots,n\},\, i \text{ odd}} \big[m_{15} - m_{16} + m_1 - m_2 \big] \\
& + \sum_{i \in \{1,\ldots,n-2\},\, i \text{ odd}} \big[m_{12} + m_{10} \big] + M(B) + M(BC_1 \cdots C_n) \\
& + \sum_{i \in \{3,\ldots,n\},\, i \text{ odd}} \big[-m_{11} - m_9 \big] - M(BC_1 \cdots C_n) - M(B) \\
& + \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[m_7 + m_4 - m_8 - m_3 + m_{14} + m_5 - m_{13} - m_6 \big] \quad \Leftrightarrow \\
0 \;=\; & \sum_{i \in \{1,\ldots,n\},\, i \text{ odd}} \big[m_{15} + m_{12} - m_{16} - m_{11} + m_{10} + m_1 - m_9 - m_2 \big] \\
& + \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[m_7 + m_4 - m_8 - m_3 + m_{14} + m_5 - m_{13} - m_6 \big]
\end{aligned}
$$

which is true if and only if

$$
\bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ odd}}} \big[\langle A, C_i \mid B\, C_{i+1} \cdots C_n \rangle \;\wedge\; \langle B, C_i \mid C_1 \cdots C_{i-1} \rangle \big] \;\wedge \bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ even}}} \big[\langle A, C_i \mid C_{i+1} \cdots C_n \rangle \;\wedge\; \langle B, C_i \mid A\, C_1 \cdots C_{i-1} \rangle \big]
$$

are valid independence statements. □

Proposition 2

We have that

$$
\bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ odd}}} \big[\langle A, C_i \mid C_1 \cdots C_{i-1} \rangle \;\wedge\; \langle B, C_i \mid A\, C_{i+1} \cdots C_n \rangle \big] \;\wedge \bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ even}}} \big[\langle A, C_i \mid B\, C_{i+1} \cdots C_n \rangle \;\wedge\; \langle B, C_i \mid C_1 \cdots C_{i-1} \rangle \big]
$$

are valid independence statements if and only if:

$$
\begin{aligned}
0 \;=\; & \sum_{i \in \{1,\ldots,n\},\, i \text{ odd}} \big[m_6 + m_1 - m_5 - m_2 + m_{15} + m_8 - m_{16} - m_7 \big] \\
& + \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[m_{15} + m_{12} - m_{16} - m_{11} + m_{10} + m_1 - m_9 - m_2 \big] \quad \Leftrightarrow
\end{aligned}
$$

$$\begin{aligned}
0 \;=\; & \sum_{i \in \{1,\ldots,n-2\},\, i \text{ odd}} \big[\, m_6 - m_2 + m_8 - m_{16} \,\big] \\
& \quad + M(AC_1 \cdots C_n) - M(C_1 \cdots C_n) + M(A) - M(AB) \\
& + \sum_{i \in \{3,\ldots,n\},\, i \text{ odd}} \big[\, m_1 - m_5 + m_{15} - m_7 \,\big] \\
& \quad + M(\varnothing) - M(A) + M(ABC_1 \cdots C_n) - M(AC_1 \cdots C_n) \\
& + \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[\, m_{15} + m_{12} - m_{16} - m_{11} + m_{10} + m_1 - m_9 - m_2 \,\big]
\end{aligned}$$

We observe that

$$\sum_{i \in \{3,\ldots,n\},\, i \text{ odd}} \big[\, m_1 + m_{15} \,\big] \;+\; \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[- m_2 - m_{16} \,\big] \;=\; 0,$$

$$\sum_{i \in \{1,\ldots,n-2\},\, i \text{ odd}} \big[- m_2 - m_{16} \,\big] \;+\; \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[\, m_1 + m_{15} \,\big] \;=\; 0$$

and thus find

$$\begin{aligned}
0 \;=\; & \sum_{i \in \{1,\ldots,n-2\},\, i \text{ odd}} \big[\, m_6 + m_8 \,\big] \\
& + \sum_{i \in \{3,\ldots,n\},\, i \text{ odd}} \big[- m_5 - m_7 \,\big] \\
& + \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[\, m_{12} - m_{11} + m_{10} - m_9 \,\big] \\
& + M(\varnothing) - M(C_1 \cdots C_n) + M(ABC_1 \cdots C_n) - M(AB)
\end{aligned}$$

Since

$$\begin{aligned}
\sum_{i \in \{1,\ldots,n-2\},\, i \text{ odd}} \big[\, m_6 + m_8 \,\big] &= \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[\, m_5 + m_7 \,\big] \\
\sum_{i \in \{3,\ldots,n\},\, i \text{ odd}} \big[- m_5 - m_7 \,\big] &= \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[- m_6 - m_8 \,\big] \\
\sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[\, m_{12} + m_{10} \,\big] &= \sum_{i \in \{3,\ldots,n\},\, i \text{ odd}} \big[\, m_{11} + m_9 \,\big]
\end{aligned}$$

$$\sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[-m_{11} - m_9 \big] \quad = \sum_{i \in \{1,\ldots,n-2\},\, i \text{ odd}} \big[-m_{12} - m_{10} \big]$$

and moreover

$$\sum_{i \in \{1,\ldots,n-2\},\, i \text{ odd}} \big[m_{14} + m_4 \big] + \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[-m_{13} - m_3 \big] \quad = \quad 0$$

$$\sum_{i \in \{3,\ldots,n\},\, i \text{ odd}} \big[-m_{13} - m_3 \big] + \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[m_{14} + m_4 \big] \quad = \quad 0,$$

$$M(B) - M(B) - M(BC_1 \cdots C_n) + M(BC_1 \cdots C_n) \quad = \quad 0$$

we have that

$$\begin{aligned}
0 \;=\; & \sum_{i \in \{1,\ldots,n-2\},\, i \text{ odd}} \big[m_{14} - m_{10} + m_4 - m_{12} \big] \\
& + M(ABC_1 \cdots C_n) - M(BC_1 \cdots C_n) + M(\varnothing) - M(B) \\
+ & \sum_{i \in \{3,\ldots,n\},\, i \text{ odd}} \big[m_9 - m_{13} + m_{11} - m_3 \big] \\
& + M(B) - M(AB) + M(BC_1 \cdots C_n) - M(C_1 \cdots C_n) \\
+ & \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[m_7 + m_4 - m_8 - m_3 + m_{14} + m_5 - m_{13} - m_6 \big] \quad \Leftrightarrow
\end{aligned}$$

$$\begin{aligned}
0 \;=\; & \sum_{i \in \{1,\ldots,n\},\, i \text{ odd}} \big[m_{14} + m_9 - m_{13} - m_{10} + m_{11} + m_4 - m_{12} - m_3 \big] \\
+ & \sum_{i \in \{1,\ldots,n\},\, i \text{ even}} \big[m_7 + m_4 - m_8 - m_3 + m_{14} + m_5 - m_{13} - m_6 \big]
\end{aligned}$$

which is true if and only if

$$\bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ odd}}} \big[\langle A, C_i \mid B\, C_1 \cdots C_{i-1} \rangle \;\wedge\; \langle B, C_i \mid C_{i+1} \cdots C_n \rangle \big] \;\wedge \bigwedge_{\substack{i \in \{1,\ldots,n\},\\ i \text{ even}}} \big[\langle A, C_i \mid C_{i+1} \cdots C_n \rangle \;\wedge\; \langle B, C_i \mid A\, C_1 \cdots C_{i-1} \rangle \big]$$

are valid independence statements. □

Hybrid evaluation of the industrial global impact on Mexican aquifers under uncertain criteria evaluations

Horacio Flores Casamayor[1], Silvia Carpitella[2], Joaquín Izquierdo[3], Jesús Mora-Rodríguez[1], and Xitlali Delgado-Galván[1]

[1]Geomatics and Hydraulics Engineering Department, Universidad de Guanajuato, Av. Juárez 77, 36000, Guanajuato, México;
{h.florescasamayor,jesusmora,xdelgado}@ugto.mx
[2]Department of Decision-Making Theory, Institute of Information Theory and Automation, Czech Academy of Sciences, Pod Vodárenskou Věží 4, 18208, Prague, Czech Republic;
carpitella@utia.cas.cz
[3]Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Camino de Vera s/n, 46022, València, España;
jizquier@upv.es

Abstract

The present paper proposes an integrated methodological approach to address the problem of managing five aquifers of Guanajuato state, Mexico, according to such relevant criteria as environmental, social, economic and hydrological aspects. The goal of this research consists in formalizing a structured framework to first evaluate the various degrees of importance of criteria and to secondly get a classification of aquifers by minimizing uncertainty of evaluations. To such an aim, the Analytic Hierarchy Process (AHP) is used for calculating the vector of criteria weights, while the Fuzzy Logic (FL) theory supports in deriving quantitative evaluations of aquifers under each selected criterion. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is then proposed to formalize the final ranking of aquifers, something that will be helpful to understand which alternative matches all the differently weighted criteria in the most suitable way at a practical level. In such a way, getting a comprehensive and strategic overview about the problem of interest will be possible.

1 Introduction

Guanajuato state is located in northern Mexico, and a part of its territory is included within the region known as "El Bajío" which, due to its climatic and geographical conditions, has a rainfall regime whose average annual precipitation is lower than the national average (Figure 1). The region has scarce surface water sources and is susceptible to suffer periods of drought. Groundwater is the main water supply for different productive and domestic uses, which, due to the extraction and management policies implemented over time, presents various degrees of affectation due to overexploitation and pollution.

Problems of water scarcity are not new in this region, since it has suffered important allocation problems; ownership and use of waterways have been a cause of frequent disputes since the colonial period (Seligmann, 1988). The surface water has had important restrictions since 1931 (DOF, 1931), whose validity has recently been highlighted on April 8, 2014 (DOF, 2014). Use of groundwater restrictions date back to 1948 (DOF, 1948) and groundwater still continues nowadays to be restricted, giving priority to domestic use in any case. Due to the above, a new economic policy in the state of Guanajuato has stimulated the establishment of various industries. For example, those related to the automotive sector. Also, the change in the types of crops with the aim of increasing economic and social development by allocating the water of the area to activities considered more productive. And these actions have been taken without apparently measuring the costs in the ecological sustainability that this policy can cause in the whole area.

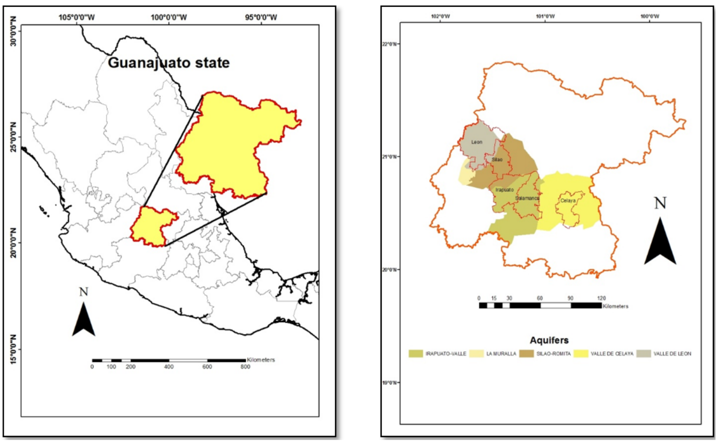


Figure 1: Localization of Guanajuato state, aquifers and municipalities of study

The present work selected five municipalities in the state of Guanajuato, namely Celaya, Irapuato, León, Salamanca and Silao, which are among the main recipients of investment in the state. They have industrial and agricultural sectors with plenty of economic weight, and currently have strong problems in hydrological matters due to scarcity, over exploitation, increased water demand and pollution of its main source of water, aquifers.

The main objective of this work is to analyze the elements that are important to evaluate water management, and the evolution of each one of these elements considering the available data, in order to determine if the water policy is achieving good results. The Analytic Hierarchy Process (AHP) and the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) were selected along with the Fuzzy Logic (FL) theory.

2 Methodological details

2.1 Basics of AHP and FL for uncertain criteria evaluations

The AHP methodology was developed by Saaty as a way to study and solve complex problems by dividing them into smaller components, constructing a tree of decision where the interrelation between elements is established visually; later, comparison matrices are formed (Saaty, 1977, 1987). The tree of decision is integrated by the problem or objective to be solved, which is placed in the upper part. Immediately below, there are the criteria, which are the main issues or elements that constitute the problem and, if it is convenient to be more specific, the criteria can be subdivided into subcriteria. Finally, at the bottom of the tree, the alternatives, which are the proposed options for solving the problem.

To solve the problem, AHP takes the opinion, experience and way of thinking of people with knowledge in the problem addressed to obtain, through mathematical processes, the most viable option to solve it. Now, in order to compare the opinions expressed by the experts, different ratio scales are used, which can be numerical, verbal or graphic. In this case, the numerical scale of values designed by Saaty (1977), which covers values from 1 to 9 to assign the importance to the options, where 1 means equal importance and 9 extreme importance; a detailed analysis of this scale is presented in (Ishizaka and Nemery, 2013).

The opinion is addressed by pairwise comparing elements, and comparisons are summarized into so-called pairwise comparison matrices. The Perron eigenvector, $W = [w_1, w_2, w_3, \ldots, w_n]$, of any of these positive matrices, gives the vector of priorities, whose components, w_n, indicate the weights or importance of the considered elements, to reach the best solution. Over time, AHP has been applied in various fields of science, technology, industry, among others, alone and in combination with other methodologies. The literature shows proposals of management of water supply problems by means of a multi-criteria point of view (Ilaya-Ayza et al., 2017; Kourgialas et al., 2019; Singh et al., 2017), demonstrating its efficacy. It is also fundamental to take into account such aspects as uncertainty affecting decisions in water management (Höllermann and Evers, 2017) and differences that may exist among opinions given by decision-makers (Tembata and Takeuchi, 2018).

In general, as asserted by Yager (2018), many modern technological tasks make use of multi-criteria methods, and evaluation criteria are usually categorized to express information about their mutual importance. Such authors as Safarzadeh et al. (2018) consider multi-criteria decision-making methods to be among the most helpful rational mathematical approaches of the last decades for selecting appropriate alternatives. Moreover, their combinations with other methodologies increases their accuracy (Che et al., 2010; Ramanathan, 2006).

FL theory, developed by Zadeh, was applied to determine the degree of belonging of an element within a set of elements (Zadeh, 1965). This is done using linguistic variables such as “a lot”, “very”, “a little”, which are defined based on the opinion of experts. Therefore, a proposition may be partially true or false and allows to consider if an element belongs to a set with a certain degree of membership. This degree is expressed with a numeric value in the interval (0, 1), which allows one to simulate the human way of reasoning. To apply the methodology, it is necessary to follow the next steps (Mahabir et al., 2003).

a) *Defining a set of variables and assigning a membership function defining the degree of belonging that each variable has in a group, indicated usually with a linguistic term.*

b) *Defining rules to relate each variable to its membership function with the obtained result, usually through a series of IF-THEN rules, IF representing a condition and THEN a conclusion.*

c) *Evaluating statements or rules mathematically and applying defuzzification to get crisp results.*

There are various methodologies for defuzzification, such as the Center of Area (COA), Center of Gravity (COG), and Mean of Maxima (MeOA). In our case, the trapezoidal membership function was used, being the function that best adjusted to the behavior of the variables. Values are given for the corresponding intervals by Functions (1) and (2):

$$\mu_A(x) = \begin{cases} & 0 \leq x \leq a \\ \frac{x-a}{b-a} & a < x \leq b; \\ & x > b \end{cases} \tag{1}$$

$$\mu_A(x) = \begin{cases} & 0 \leq x \leq a \\ \frac{b-x}{b-a} & a < x \leq b. \\ & x > b \end{cases} \tag{2}$$

In the above, a indicates the lower limit; b indicates the upper limit; x is the value to estimate; $\mu_A(x)$ is the membership function for a fuzzy set A on the universe of discourse X, and is defined as $\mu_A(x) : X \rightarrow [0, 1]$. FL has been adapted over time to the control of processes related to production systems (McBratney and Odeh, 1997; Kommadath et al., 2012), management and treatment of water (Boiocchi et al., 2016; Mahabir et al., 2003), in transport systems (Rajak et al., 2016), agricultural production (Kavdir and Guyer, 2004; Center and Verma, 1998), mining (Hüllermeier, 2011), among many other areas which have been susceptible to be automated. Likewise, FL has been applied in conjunction with other analysis methodologies to increase the effectiveness of both as in the case of the method TOPSIS (Sanghvi et al., 2021).

2.2 TOPSIS procedure for ranking decision-making alternatives

The TOPSIS technique, originally developed by Hwang and Yoon (1981) with further developments by Yoon (1987) and Hwang et al. (1993), is an established multi-criteria decision-making method useful to rank alternatives representing potential solutions of a given decision-making problems in many application areas (Ouenniche et al., 2018; Nilashi et al., 2019; Meniz, 2021). The method is based on the concept of calculating distances between each alternative and two ideal points, namely, a positive ideal solution and a negative ideal solution. In such a way, the alternative(s) occupying the first position(s) in the final ranking will be that one(s) closest to the positive ideal solution and farthest from the negative ideal solution. TOPSIS-based approaches have been proposed within the context of water quality evaluation (Li et al., 2018) also in integration with the AHP technique (Xu et al., 2016; Zyoud et al., 2016; Fu et al., 2013). Regarding the practical application of the methodology, it is necessary to preliminary collect and organize the following input data: quantitative evaluations of alternatives under each criterion, vector of criteria weights (reflecting the mutual importance of the considered aspects), preference direction of criteria (establishing if criteria need to be maximized or minimized). Once weights have been assigned to criteria and established their preference directions, alternatives have to be ranked by implementing the phases described in such works as (De Anchieta et al., 2021). In particular, the following stages need to be implemented.

- *Compiling the input decision-making matrix by collecting the assessments g_{ij} related to each alternative i under each criterion j taken into account for the evaluation.*
- *Computing the weighted and normalised decision-making matrix, for which the generic element u_{ij} can be calculated as follows:*
$$u_{ij} = w_j \times z_{ij}, \forall i, \forall j; \tag{3}$$
where w_j represents the weight of criterion j, and z_{ij} the score of the generic alternative i under the mentioned criterion j, normalised by means of the following operation:
$$z_{ij} = \frac{g_{ij}}{\sqrt{\sum_{i=1}^{n} g_{ij}^2}}. \tag{4}$$
- *Identifying two points representing ideal solutions, namely the positive ideal solution A^* and the negative ideal solution A^-, by means of the following equations:*
$$A^* = (u_1^*, \ldots, u_k^*) = \{(\max_i u_{ij} | j \in I^{'}), (\min_i u_{ij} | j \in I^{''})\}; \tag{5}$$
$$A^- = (u_1^-, \ldots, u_k^-) = \{(\min_i u_{ij} | j \in I^{'}), (\max_i u_{ij} | j \in I^{''})\}; \tag{6}$$
where $I^{'}$ and $I^{''}$ are the sets of criteria to be, respectively, maximised and minimised.
- *Calculating S^* as the distance from each alternative i to the positive ideal solution A^* and S^- as the distance from each alternative i to the negative ideal solution A^- as follows:*
$$S^* = \sqrt{\sum_{j=1}^{k} (u_{ij} - u_{ij}^*)}, i = 1, \ldots, n; \tag{7}$$
$$S^- = \sqrt{\sum_{j=1}^{k} (u_{ij} - u_{ij}^-)}, i = 1, \ldots, n. \tag{8}$$
- *Computing, for each solution i, the closeness coefficient C_i^* representing how alternative i performs with respect to the previously calculated ideal solutions:*
$$C_i^* = \frac{S^-}{S^- + S^*}, 0 < C_i^* < 1, \forall i. \tag{9}$$
- *Ranking the available decision-making alternatives by ordering the calculated closeness coefficients in a decreasing way. This means that, when referring to two generic alternatives i and z, if $C_i^* \geq C_z^*$ solution i should be preferred to solution z.*

3 Application and results

AHP methodology has been carried out by involving 48 specialists in the water sector belonging to agencies such as the State Water Commission (CEA), Groundwater Technical Committees (COTAS), the Municipal Water Utilities and the University of Guanajuato (UG). The main objective was to define the importance that each criteria and subcriteria should have on the water management for the study area, considering the state policy to receive new companies, mostly industries. Once verified that the matrices met the conditions to use them, the results obtained for the different elements are shown below, close together with the tree of decision that was defined (Figure 2), and both are widely explained in Flores Casamayor et al. (2018). We can observe as, according to the opinions expressed by the involved experts, the criterion referring to “Hydrological Aspects” has been considered as the most important.

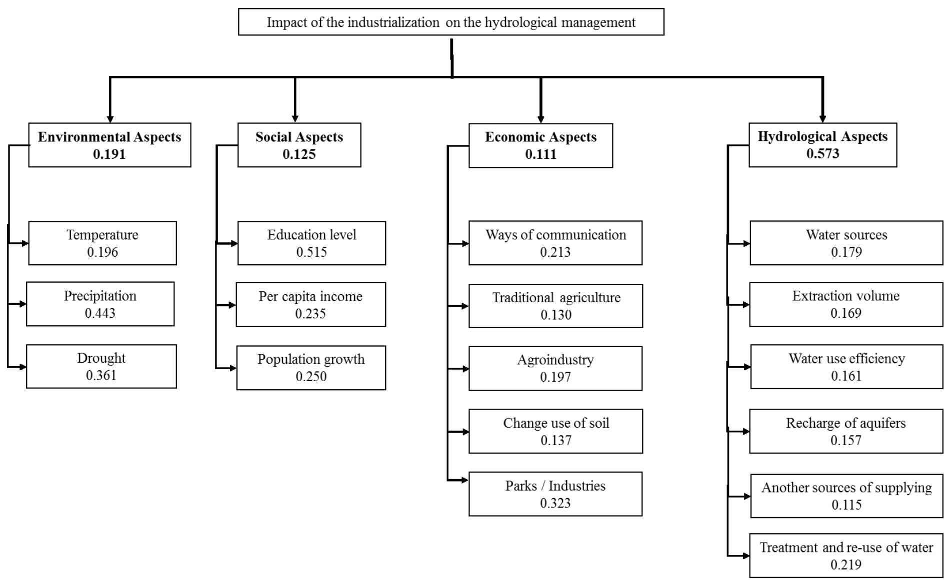


Figure 2: Tree of decisions and AHP results

This is congruent if we consider the current problem that the municipalities of the study present in the aquifers where they extract all the water to cover the requirements of their population and its economy. With relation to this criterion, subcriterion of “Treatment and reuse of water” was considered as the most important, which indicates that the water discarded by the various users is likely to be used in those activities that do not necessarily require potable water to be taken as a measure to reduce overexploitation of water sources. The order of importance in the rest of the criteria was: Environmental, Social and Economic Aspects; in the case of subcriteria, the most important were, respectively, Precipitation, Education level and Parks/Industries.

The FL methodology was applied to the whole set of subcriteria defined in AHP, with the aim to analyze the evolution of each parameter, by preferably considering the range of ten years. To get a broad explanation about the FL process, such as the type of graphic used in each case, the source of the values of the different parameters, the value of FL assigned to them according the performed methodology, and so on, it is necessary to consult Flores Casamayor et al. (2021). Next Table 1 shows a summary of the results of both methodologies, where we can see the relations and coincidences between the values calculated in all cases. In the case of "Treatment of water" and "Reuse of water" parameters, it is necessary to clarify that their values are equal because originally both of them were considered as a single parameter in the AHP (see Figure 2).

However, data obtained from official sources and used in the analysis of FL were provided as two separated parameters; so to keep the consistency it was decided to divide the AHP original value of 0.219, getting the value of 0.1095 reported in Table 1. In the beginning, the value 1 was assigned to the subcriterion "Water sources" to the municipality of León. Since 1995 El Zapotillo dam project was started to solve water supply problems in Jalisco y Guanajuato, including León (Briseño Ramírez, 2021). However, El Zapotillo did not advance in its construction due to various legal demands made by different civil organizations, issued on the great environmental, social, and economic impacts that such work would cause in the area where it is located. The project was finally resumed in November 2021, but no water resources were considered for León (CONAGUA, 2021). Then, the value in Table 1 was changed to zero.

				FL values				
Criteria AHP	**Values**	**Subcriteria AHP**	**Values**	**Celaya**	**Irapuato**	**León**	**Salamanca**	**Silao**
Environmental	0.192	Temperature	0.196	0.000	1.000	0.533	0.617	0.350
		Precipitation	0.443	1.000	1.000	1.000	0.912	0.564
		Drought	0.361	0.742	0.723	0.754	0.729	0.723
Social	0.125	Education level	0.515	0.520	0.467	0.487	0.413	0.278
		Per capita income	0.235	0.000	0.000	0.000	0.000	0.000
		Population growth	0.250	1.000	0.800	0.650	1	0.700
Economic	0.111	Ways of communic.	0.213	0.708	0.622	0.667	0.868	0.667
		Trad. agriculture	0.130	0.698	0.500	1.000	0.750	0.250
		Agroindustry	0.197	0.995	0.858	0.917	0.750	0.726
		Change use of soil	0.137	0.714	0.571	0.571	0.000	0.286
		Parks/Industries	0.323	0.693	0.663	0.286	0.773	0.767
Hydrological	0.573	Water sources	0.179	0.353	0.000	0.000	0.000	0.000
		Extraction volume	0.169	0.000	0.492	0.453	0.492	0.000
		Water use efficiency	0.161	0.419	0.419	0.419	0.419	0.419
		Recharge of aquifers	0.157	0.000	0.000	0.468	0.000	0.000
		Other sources	0.115	0.061	0.063	0.077	0.097	0.132
		Treatment of water	0.1095	1.000	0.498	1.000	1.000	1.000
		Reuse of water	0.1095	0.494	0.101	0.861	0.007	0.012

Table 1: Comparison of results AHP - FL

A feasibility criterion has been designed in order to interpret if the industries established in the area of study would have any positive impact on the defined criteria and subcriteria defined. The feasibility criterion was defined as reported in Table 2.

Next table shows the Industrial global impact to the municipalities under study. First, it was calculated by multiplying the AHP and FL values obtained for each subcriterion, adding the parameters that integrate each criterion and achieving a total result in each case for all the municipalities. Later, these values obtained for the subcriteria were multiplied for the corresponding AHP criterion value, added and finally, the Industrial global impact was obtained (Table 3).

$0 < X \leq 0.250$: **Very Unfavorable Trend** (VUT). The objective is not met and there is considerable deterioration or impact on various parameters. The establishment of new industries is not recommended until the parameters are balanced or consistent improvements are observed. The impact of industries on improving the water situation is irrelevant (they may be a factor that aggravates the situation).
$0.250 < X \leq 0.50$: **Unfavorable Trend** (UT). The objectives are not met and different levels of deterioration are presented in various aspects. It is not recommended to accept the arrival of new industries, and in those already established there must be strict supervision in their water consumption and waste discharge. Their contribution to solving water problems is marginal or unimportant.
$0.50 < X \leq 0.75$: **Positive but Insufficient Progress** (PIP). The objectives are being met, but there are still aspects that require special attention. The arrival of new industries can be considered, but with strict supervision of their water consumption and waste dumping, among other limitations. Their contribution to solving water problems has some relevance.
$0.75 < X \leq 1$: **Positive Trend** (PT). The goals are being reached and the parameters are in balance or close to being therein. New industries can be allowed to arrive if they meet the relevant standards and have resource-efficient technologies and infrastructure for the treatment of their wastewater. The contribution of present companies to solving water-related problems is remarkable.

Table 2: Feasibility criterion

For a wider explanation about the process that was followed to get the values shown in the table above, readers are encouraged to consult Flores Casamayor et al. (2021).

Criteria	AHP Values	Celaya	Irapuato	León	Salamanca	Silao
Environmental	0.191	0.711	0.900	0.820	0.788	0.579
Social	0.125	0.518	0.441	0.414	0.463	0.318
Economic	0.111	0.781	0.683	0.649	0.697	0.565
Hydrological	0.573	0.301	0.223	0.430	0.272	0.193
Industrial global impact		0.460	0.430	0.530	0.440	0.320
Feasibility criterion		UT	UT	PIP	UT	UT

Table 3: Industrial global impact on the municipalities of study

The TOPSIS methodology has been lastly applied to the dataset of Table 3 to obtain the ranking of the analysed aquifers, formalised in Table 4. The application has been led by normalizing subcriteria weights and defining their preference directions. Some preliminary considerations about preference directions of subcriteria need to be underlined.

- *Hydrological aspects: six out of seven subcriteria related to the hydrological aspects have been maximized. They are treatment and reuse of water, water sources, water use efficiency, recharge of aquifers and other supply sources, all of them representing measures aimed at reducing pressures on water sources. On the contrary, the subcriterion of extraction volume has been minimised because it represents the main problem with water sources in the analysed municipalities.*

- *Environmental aspects: precipitation subcriterion needs to be maximized, since higher associated values associated with such aspect have to be preferred in terms of aquifers management, while drought and temperature subcriteria need to be minimised.*

- *Social aspects: the proposal consists in maximizing subcriteria referring to the educational level and per capita income, that are the parameters for which an increase would be beneficial, while simultaneously minimizing the subcriterion of population growth, whose increase would be directly related with an increase of cost.*

- *Economic aspects: all the subcriteria belonging to this category should be maximized. In particular, subcriteria representing traditional agriculture and agro-industry currently represent the main sources of water consumption for the municipalities under study, occupying the vast majority of the available surface. Let us note that the maximization of subcriteria does not refer to the need of expanding the occupied surface but to their efficiency, so that reducing their impact on the geographical area of reference will be possible.*

The final ranking of alternatives confirms that the León aquifer occupies the first position. Results reported in Table 4 also confirm the considerations previously expressed when evaluating the industrial global impact in the municipalities of study (Table 3).

Aquifers	S_i^*	S_i^-	C_i*	Ranking position
Celaya	0.082	0.097	0.542	2^{nd}
Irapuato	0.109	0.046	0.298	5^{th}
León	0.060	0.084	0.583	1^{st}
Salamanca	0.107	0.049	0.313	4^{th}
Silao	0.101	0.083	0.449	3^{rd}

Table 4: Final ranking obtained by means of the TOPSIS technique

4 Discussion and conclusions

The region of El Bajío has been affected by problems of water management for many years. There is a close relationship between the opinions issued by the experts in the AHP, regarding the priority of water aspects, and the current situation of the aquifers that supply the municipalities. This also highlights the importance given to the treatment and reuse of water as a way to reduce water extraction from aquifers, where it surpasses the effort that needs to be made to better take advantage of the available treated water.

Considering the environmental aspects as a whole, despite the AHP results indicate greater importance to precipitation - supported in part by the values available from official sources - further observations of the parameters are still needed to get a clearer picture of the trend that will prevail in each municipality. Indeed, climatic changes will imply significant oscillations on the values of these parameters in medium and long term.

Considering the economic aspects, the most important criterion according to AHP was the industrial activity, which makes sense because of the current development policy implemented in the municipalities. This is supported by the number of companies arriving each year; in turn, to facilitate the arrival of industries, the construction of highways and roads in municipalities was promoted. So, if we take all the values obtained in both methodologies and the index of industrial global impact, this economic policies have provided results and benefits more than questionable in the municipalities, and it is necessary to re-think the type of development to follow in the municipalities of study, and in the region of El Bajío. El Zapotillo will not eventually allocate water for the water supply in León. The impact in the Feasibility criterion is remarkable, because in this case its value will be 0.530, which is catalogued as “Positive but insufficient progress” (PIP), closer to the rest of the municipalities. The final TOPSIS application confirms these results, since the León aquifer occupies the first position of the ranking in both situations, that is without the implementation of the project El Zapotillo for León.

Acknowledgements

The research is financially supported by the Czech Science Foundation, Grant No. 19-06569S.

References

R. Boiocchi, K. V. Gernaey, and G. Sin. Control of wastewater N_2O emissions by balancing the microbial communities using a fuzzy-logic approach. *IFAC-PapersOnLine*, 49(7): 1157–1162, 2016.

H. Briseño Ramírez. The water supply proyect "El zapotillo": facts and perceptions. *Revista Economía y Política*, (33):34–47, 2021.

B. Center and B. P. Verma. Fuzzy logic for biological and agricultural systems. In *Artificial Intelligence for Biology and Agriculture*, pages 213–225. Springer, 1998.

Z. Che, H. Wang, and C.-L. Chuang. A fuzzy AHP and DEA approach for making bank loan decisions for small and medium enterprises in Taiwan. *Expert Systems with Applications*, 37(10):7189–7199, 2010.

CONAGUA. Comunicado de Prensa No. 900-21. https://www.gob.mx/cms/uploads/attachment/file/681151/Comunicado_de_Prensa_Presa_No._900-21.pdf, 2021.

T. F. F. De Anchieta, S. A. Santos, B. M. Brentan, S. Carpitella, and J. Izquierdo. Managing expert knowledge in water network expansion project implementation. *IFAC-PapersOnLine*, 54(17):36–40, 2021.

DOF. Acuerdo que establece veda sobre concesión de aguas del río Lerma y sus afluentes, en los Estados de México, Michoacán, Guanajuato y Jalisco. *Diario Oficial de la Federación (DOF)*, 1931.

DOF. Decreto que establece por tiempo indefinido, veda para la construcción o ampliación de las obras de alumbramiento de aguas del subsuelo, que comprenderá la zona perimetral que ocupaba el Distrito de León, Gto. *Diario Oficial de la Federación (DOF)*, 1948.

DOF. Decreto por el que por causas de interés público se suprimen las vedas existentes en la subregión hidrológica Lerma-Chapala, y se establece zona de veda en las 19 cuencas hidrológicas que comprende dicha subregión hidrológica. *Diario Oficial de la Federación (DOF)*, 2014.

H. Flores Casamayor, J. L. Morales Martínez, J. Mora-Rodríguez, G. Carreño, and X. Delgado-Galván. Management priorities for aquifers in el bajío in guanajuato state, mexico. *Water Policy*, 20(6):1161–1175, 2018.

H. Flores Casamayor, J. L. Morales Martínez, J. Mora-Rodríguez, and X. Delgado-Galván. Assessing industrial impact on water sustainability in el bajío, guanajuato state, mexico. *Sustainability*, 13(11):6161, 2021.

H.-Z. Fu, M.-H. Wang, and Y.-S. Ho. Mapping of drinking water research: A bibliometric analysis of research output during 1992–2011. *Science of the Total Environment*, 443: 757–765, 2013.

B. Höllermann and M. Evers. Perception and handling of uncertainties in water management—a study of practitioners' and scientists' perspectives on uncertainty in their daily decision-making. *Environmental Science & Policy*, 71:9–18, 2017.

E. Hüllermeier. Fuzzy machine learning and data mininga. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(4):269–283, 2011.

C. Hwang and K. Yoon. Multiple Attribute Decision Making. Methods and Applications. In *Springer Verlag, Berlin Heidelberg New York*. 1981.

C.-L. Hwang, Y.-J. Lai, and T.-Y. Liu. A new approach for multiple objective decision making. *Computers & operations research*, 20(8):889–899, 1993.

A. E. Ilaya-Ayza, J. Benítez, J. Izquierdo, and R. Pérez-García. Multi-criteria optimization of supply schedules in intermittent water supply systems. *Journal of Computational and Applied Mathematics*, 309:695–703, 2017.

A. Ishizaka and P. Nemery. *Multi-criteria decision analysis: methods and software*. John Wiley & Sons, 2013.

I. Kavdir and D. E. Guyer. Apple grading using fuzzy logic. *Turkish Journal of Agriculture and Forestry*, 27(6):375–382, 2004.

B. Kommadath, R. Sarkar, and B. Rath. A fuzzy logic based approach to assess sustainable development of the mining and minerals sector. *Sustainable Development*, 20(6): 386–399, 2012.

N. N. Kourgialas, G. C. Koubouris, and Z. Dokou. Optimal irrigation planning for addressing current or future water scarcity in Mediterranean tree crops. *Science of The Total Environment*, 654:616–632, 2019.

Z. Li, T. Yang, C.-S. Huang, C.-Y. Xu, Q. Shao, P. Shi, X. Wang, and T. Cui. An improved approach for water quality evaluation: TOPSIS-based informative weighting and ranking (TIWR) approach. *Ecological Indicators*, 89:356–364, 2018.

C. Mahabir, F. Hicks, and A. R. Fayek. Application of fuzzy logic to forecast seasonal runoff. *Hydrological processes*, 17(18):3749–3762, 2003.

A. B. McBratney and I. O. Odeh. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma*, 77(2-4):85–113, 1997.

B. Meniz. An advanced TOPSIS method with new fuzzy metric based on interval type-2 fuzzy sets. *Expert Systems with Applications*, 186:115770, 2021.

M. Nilashi, S. Samad, A. A. Manaf, H. Ahmadi, T. A. Rashid, A. Munshi, W. Almukadi, O. Ibrahim, and O. H. Ahmed. Factors influencing medical tourism adoption in malaysia: A dematel-fuzzy topsis approach. *Computers & Industrial Engineering*, 137:106005, 2019.

J. Ouenniche, B. Pérez-Gladish, and K. Bouslah. An out-of-sample framework for TOPSIS-based classifiers with application in bankruptcy prediction. *Technological Forecasting and Social Change*, 131:111–116, 2018.

S. Rajak, P. Parthiban, and R. Dhanalakshmi. Sustainable transportation systems performance evaluation using fuzzy logic. *Ecological Indicators*, 71:503–513, 2016.

R. Ramanathan. Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process. *Computers & Operations Research*, 33(5):1289–1307, 2006.

R. W. Saaty. The analytic hierarchy process—what it is and how it is used. *Mathematical modelling*, 9(3-5):161–176, 1987.

T. L. Saaty. A scaling method for priorities in hierarchical structures. *Journal of mathematical psychology*, 15(3):234–281, 1977.

S. Safarzadeh, S. Khansefid, and M. Rasti-Barzoki. A group multi-criteria decision-making based on best-worst method. *Computers & Industrial Engineering*, 126:111–121, 2018.

N. Sanghvi, D. Vora, J. Patel, and A. Malik. Optimization of end milling of Inconel 825 with coated tool: A mathematical comparison between GRA, TOPSIS and Fuzzy Logic methods. *Materials Today: Proceedings*, 38:2301–2309, 2021.

L. J. Seligmann. Irrigation in the Bajio region of colonial Mexico (Review). *Technology and Culture*, 9(3):679–682, 1988.

L. K. Singh, M. K. Jha, and V. Chowdary. Multi-criteria analysis and GIS modeling for identifying prospective water harvesting and artificial recharge sites for sustainable water supply. *Journal of Cleaner Production*, 142:1436–1456, 2017.

K. Tembata and K. Takeuchi. Collective decision making under drought: An empirical study of water resource management in Japan. *Water resources and economics*, 22: 19–31, 2018.

J. Xu, P. Feng, and P. Yang. Research of development strategy on China's rural drinking water supply based on SWOT–TOPSIS method combined with AHP-Entropy: a case in Hebei Province. *Environmental Earth Sciences*, 75(1):1–11, 2016.

R. R. Yager. Categorization in multi-criteria decision making. *Information Sciences*, 460: 416–423, 2018.

K. Yoon. A reconciliation among discrete compromise solutions. *Journal of the Operational Research Society*, 38(3):277–286, 1987.

L. A. Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

S. H. Zyoud, S. H. Zyoud, S. W. Al-Jabi, W. M. Sweileh, and R. Awang. Contribution of Arab countries to pharmaceutical wastewater literature: a bibliometric and comparative analysis of research output. *Annals of occupational and environmental medicine*, 28(1):1–12, 2016.

Independence and correlation for coherent lower conditional probabilities

Giulianella Coletti[1], Sara Latini[2], and Davide Petturiti[3]

[1]Dept. Mathematics and Computer Science, University of Perugia
giulianella.coletti@unipg.it
[2]Dept. Information Engineering, Computer Science and Mathematics, University of L'Aquila
sara.latini@graduate.univaq.it
[3]Dept. Economics, University of Perugia
davide.petturiti@unipg.it

Abstract

Coherent lower conditional probability is the main tool in decision models representing the ambiguity aversion of the decision maker, such as in the pessimistic synthesis of probabilistic judgments, under different (also unexpected) scenarios. Aim of this paper is to put under the right perspective the concepts of independence and correlation in the framework of coherent lower conditional probability, taking suitably into account also events whose lower probability is zero or one.

1 Introduction

As far as stochastic independence and positive and negative correlation are concerned, a series of papers (see, for instance, Coletti and Scozzafava (2002a,b,c); Vantaggi (2001, 2003), Coletti et al. (2020)) has pointed out the shortcomings of classic definitions, which give rise to counterintuitive situations, in particular when the given events have probability equal to 0 or 1. The cited reinforced definitions of independence and correlation between two events agree with the classic ones and their variations when the probabilities of the relevant events are both different from 0 and 1. They are able to avoid situations, as those in the framework of classic definitions, where (contrary to intuition) logical dependence does not imply stochastic dependence, or logical constraints do not imply (positive or negative) correlation. When $E|H$ and $E|H^c$ have both 0 probability, the definitions use the comparison of the corresponding zero-layers (while, for $E|H$ and $E|H^c$ with probability 1 the zero-layers of $E^c|H$ and $E^c|H^c$ are considered). The quoted zero-layers refer to a class of unconditional probabilities representing the restriction of P to $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$.

In this paper we deal with coherent lower conditional probability, which, as is well-known, plays an important role in decision making models in presence of ambiguity. It is obvious that, if one cannot ignore the consideration of null events in the framework of coherent conditional probabilities, this is even more compelling when dealing with coherent lower conditional probabilities.

Our aim is to extend the above reinforced concepts of independence and correlation to coherent lower conditional probabilities. The definitions we propose refer to elements of the dominating class of $\underline{P}$, restricted to $\mathcal{D}$, taking the minimum in $E|H$ and $E|H^c$. So the problems related to the reinforcement of definitions of independence and correlation are present in this more general framework. In addition, the quoted problems are conjoined with difficulties arising in the context of lower probabilities that have been pointed out also with other definitions (see, for example, Couso et al. (1999)).

Since the proof of either independence or correlation relying on zero-layers can be hard, we provide some characterization theorems that only refer to conditional lower probability values on $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$.

2 The framework of reference

We refer to the framework where conditional probability is the basic concept introduced as a function of two variables (precisely an ordered pair of events) ruled by a set of axioms.

2.1 Coherent conditional probability

We consider an *event* to be any fact described by a Boolean sentence, indicating by Ω the *sure event* and using $\varnothing$ for the *impossible event*. Given an event E, we will use the notation E^* to denote either E or its contrary E^c. We recall that an *additive class* of events is a set of events closed under taking disjunction $\vee$. A *Boolean algebra* of events is an additive class which is further closed under taking the contrary $(\cdot)^c$, and hence under conjunction $\wedge$. We further take a *conditional event* $E|H$ to be an ordered pair of events E, H with $H \neq \varnothing$.

In the sequel, for any Boolean algebra $\mathcal{A}$, we write $\mathcal{A}^0$ to indicate $\mathcal{A} \setminus \{\varnothing\}$. For an arbitrary family of events $\mathcal{E} = \{E_1, \ldots, E_n\}$, we use $\mathbf{alg}(\mathcal{E})$ to denote the minimal Boolean algebra of events containing $\mathcal{E}$ and $\mathbf{add}(\mathcal{E})$ to denote the minimal additive class of events containing $\mathcal{E}$. By $\mathbf{at}(\mathcal{E})$ we indicate the finest partition of Ω contained in $\mathbf{alg}(\mathcal{E})$, in particular, the events in $\mathcal{E}$ are said to be *logically independent* if the cardinality of $\mathbf{at}(\mathcal{E})$ is 2^n.

Definition 1. *Let $\mathcal{A}$ be a Boolean algebra of events and let $\mathcal{H}$ be an additive class with $\mathcal{H} \subseteq \mathcal{A}^0$. A* ***conditional probability*** *on $\mathcal{A} \times \mathcal{H}$ is a function $P : \mathcal{A} \times \mathcal{H} \to [0, 1]$ that satisfies the following conditions:*

(i) $P(E|H) = P(E \wedge H|H)$, for every $E \in \mathcal{A}$ and $H \in \mathcal{H}$;

(ii) $P(\cdot|H)$ is a finitely additive probability on $\mathcal{A}$, for every $H \in \mathcal{H}$;

(iii) $P(E \wedge F|H) = P(E|H) \cdot P(F|E \wedge H)$, *for every* $H, E \wedge H \in \mathcal{H}$ *and* $E, F \in \mathcal{A}$.

As usual, whenever $\Omega \in \mathcal{H}$, we will write $P(E) = P(E|\Omega)$ for all $E \in \mathcal{A}$. Following Dubins (1975), we say that a conditional probability $P(\cdot|\cdot)$ is *full* on the algebra $\mathcal{A}$ if it is defined on $\mathcal{A} \times \mathcal{A}^0$. For every finite Boolean algebra of events $\mathcal{A}$ and additive class $\mathcal{H} \subseteq \mathcal{A}^0$, every conditional probability $P(\cdot|\cdot)$ on $\mathcal{A} \times \mathcal{H}$ determines a linearly ordered class $\{H_0^0, \ldots, H_0^k\}$ of decreasing elements of $\mathcal{H}$, such that:

- $H_0^0 = \bigvee_{H \in \mathcal{H}} H$;
- for $\alpha = 1, \ldots, k$, $H_0^\alpha = \bigvee\{H \in \mathcal{H} : H \subseteq H_0^{\alpha-1}, P(H|H_0^{\alpha-1}) = 0\}$.

The last event H_0^k is such that $P(H|H_0^k) > 0$ for all $H \in \mathcal{H}$ with $H \subseteq H_0^k$. The events $\{H_0, \ldots, H_k\}$ give rise to a class $\{\mathcal{I}_0, \ldots, \mathcal{I}_k\}$ of decreasing Boolean ideals of $\mathcal{A}$, where $\mathcal{I}_\alpha = \{A \in \mathcal{A} : A \subseteq H_0^\alpha\}$. In turn, the class of ideals $\{\mathcal{I}_0, \ldots, \mathcal{I}_k\}$ is associated to a class of unconditional probabilities $\{P_0, \ldots, P_k\}$ where each P_α is the restriction of $P(\cdot|H_0^\alpha)$ to $\mathcal{I}_\alpha$. In particular, each P_α is completely determined by its values on $\mathcal{I}_\alpha \cap \mathbf{at}(\mathcal{A})$ through additivity. For every event $H \in \mathcal{H}$, there is a unique index $\alpha_H \in \{0, \ldots, k\}$ such that $H \in \mathcal{I}_{\alpha_H}$ and $P_{\alpha_H}(H) > 0$. Moreover, for every $E|H \in \mathcal{A} \times \mathcal{H}$ we have that

$$P(E|H) = \frac{P_{\alpha_H}(E \wedge H)}{P_{\alpha_H}(H)}. \tag{1}$$

The class $\{P_0, \ldots, P_k\}$ is said $\mathcal{H}$*-minimal agreeing class* and allows to represent the conditional probability $P(\cdot|\cdot)$ on $\mathcal{A} \times \mathcal{H}$ through (1). In particular, a $\mathcal{A}^0$-minimal agreeing class allows to represent a full conditional probability $P(\cdot|\cdot)$ on $\mathcal{A}$ and is referred to as *complete agreeing class*.

In the sequel, we call *assessment* a function $P : \mathcal{G} \to [0,1]$ with $\mathcal{G}$ an arbitrary set of conditional events and, given $\mathcal{D} \subseteq \mathcal{G}$, $P_{|\mathcal{D}}$ stands for the *restriction* of P on $\mathcal{D}$.

Definition 2. *Let* $\mathcal{G} = \{E_i|H_i\}_{i \in I}$ *be an arbitrary family of conditional events. An assessment* $P : \mathcal{G} \to [0,1]$ *is a* **coherent conditional probability** *if there exists a conditional probability* $P' : \mathcal{A} \times \mathcal{H} \to [0,1]$, *with* $\mathcal{A} = \boldsymbol{alg}(\{E_i, H_i\}_{i \in I})$ *and* $\mathcal{H} = \boldsymbol{add}(\{H_i\}_{i \in I})$, *such that* $P'_{|\mathcal{G}} = P$.

Theorem 1. *Let* $\mathcal{G}$ *be an arbitrary family of conditional events. Then, for any function* $P : \mathcal{G} \to [0,1]$, *the following statements are equivalent:*

(i) P *is a coherent conditional probability on* $\mathcal{G}$*;*

(ii) for every finite subfamily $\mathcal{F} = \{E_1|H_1, \ldots, E_n|H_n\} \subseteq \mathcal{G}$, *there exists a* $\mathcal{H}$*-minimal agreeing class* $\{P_0, \ldots, P_k\}$ *corresponding to a conditional probability* $P'(\cdot|\cdot)$ *on* $\mathcal{A} \times \mathcal{H}$ *extending* $P_{|\mathcal{F}}$, *i.e., such that, for* $i = 1, \ldots, n$, *it holds that*

$$P(E_i|H_i) = \frac{P_{\alpha_{H_i}}(E_i \wedge H_i)}{P_{\alpha_{H_i}}(H_i)}.$$

We recall, see e.g. Coletti and Scozzafava (2002c), that coherence is a necessary and sufficient condition for the extendibility of an assessment P on any larger set $\mathcal{G}' \supset \mathcal{G}$ of conditional events. In particular, a coherent conditional probability P can be always extended (generally not in a unique way) to a full conditional probability on $\mathcal{A} = \mathbf{alg}(\{E_i, H_i\}_{i \in I})$.

To conclude our review of the setting of coherent conditional probability, we recall the concept of *zero-layer* from Coletti and Scozzafava (2002c), which naturally arises from the structure of conditional probability described in Theorem 1.

Definition 3. *Let $\mathcal{A}$ be a finite Boolean algebra of events and let $\{P_0, \ldots, P_k\}$ be a complete agreeing class on $\mathcal{A}$. For every event $H \in \mathcal{A}^0$, the* **zero-layer** *of H with respect to $\{P_0, \ldots, P_k\}$ is the non-negative number*

$$o(H) = \alpha_H,$$

where $\alpha_H \in \{0, \ldots, k\}$ is the unique index such that $H \in \mathcal{I}_{\alpha_H}$ and $P_{\alpha_H}(H) > 0$; the zero layer of the impossible event is set to $o(\varnothing) = +\infty$. For every $E|H \in \mathcal{A} \times \mathcal{A}^0$, the **zero-layer** *of $E|H$ with respect to $\{P_0, \ldots, P_k\}$ is the non-negative number*

$$o(E|H) = o(E \wedge H) - o(H).$$

The zero-layer of $E|H$ is easily seen to have a behavior that is independent of the chosen complete agreeing class in case $E \wedge H = \varnothing$ or $P(E|H) > 0$ since it reduces, respectively, to $+\infty$ and to 0. The following Theorem 2, summarizing results given in Coletti and Scozzafava (2002a) and Coletti et al. (2020), shows the robustness of the zero-layer of a conditional event with respect to the choice of the agreeing class, when E, H are logically independent events and extreme and equal probabilities are given to $E|H$ and $E|H^c$. This paves the way to using zero-layers for distinguishing between independence and correlation when extreme probability events are involved.

Theorem 2. *Let E, H be logically independent events and let P be a coherent conditional probability on a family of conditional events $\mathcal{G}$ containing the set $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$. Let $P(E|H) = P(E|H^c) = 0$ or 1. Then, if there exists a complete agreeing class $\{P_\alpha\}$ on **alg**$(\{E, H\})$ that agrees with $P_{|\mathcal{D}}$, such that one of the following conditions holds:*

(a) $o(E|H) = o(E|H^c)$ and $o(E^c|H) = o(E^c|H^c)$;

(b) $o(E|H) < [>] o(E|H^c)$;

(c) $o(E^c|H) > [<] o(E^c|H^c)$;

*then this holds true for any other complete agreeing class on **alg**$(\{E, H\})$ that agrees with $P_{|\mathcal{D}}$.*

Notice that, if $P(E|H) = P(E|H^c) = \delta \in]0, 1[$, then it trivially holds that $o(E|H) = o(E|H^c) = o(E^c|H) = o(E^c|H^c) = 0$.

2.2 Coherent lower conditional probability

Given an arbitrary set $\mathcal{G}$ of conditional events, a *coherent lower conditional probability* on $\mathcal{G}$ is a non-negative function $\underline{P}$ such that there exists a non-empty family $\mathcal{P}$ of coherent conditional probabilities on $\mathcal{G}$, said *dominating family*, whose lower envelope agrees with $\underline{P}$, that is, for every $E|H \in \mathcal{G}$,

$$\underline{P}(E|H) = \inf_{P \in \mathcal{P}} P(E|H).$$

In particular, $\mathcal{P}$ can be taken equal to the set $\mathcal{Q}$ of all dominating coherent conditional probabilities on $\mathcal{G}$

$$\mathcal{Q} = \{P : P \text{ is a coherent conditional probability on } \mathcal{G}, P \geq \underline{P}\},$$

and in this case the pointwise infima are attained (see Williams (2007)). The family $\mathcal{Q}$ will be referred to as *maximal dominating family.*

When $\mathcal{G}$ is finite, if $\underline{P}$ is a coherent lower conditional probability, then there exists a finite dominating family $\mathcal{P}$ such that

$$\underline{P}(E|H) = \min_{P \in \mathcal{P}} P(E|H),$$

whose elements are called *i-minimal* according to the following definition.

Definition 4. *Given a coherent lower conditional probability $\underline{P}$ on a finite set of conditional events $\mathcal{G} = \{E_1|H_1, \ldots, E_n|H_n\}$ and any conditional event $E_i|H_i \in \mathcal{G}$, an element P^i of the maximal dominating family $\mathcal{Q}$ such that $P^i(E_i|H_i) = \underline{P}(E_i|H_i)$ will be called* ***i*-minimal** *coherent conditional probability.*

3 Independence and correlation for coherent conditional probabilities

Inspired by what is proposed in Coletti and Scozzafava (2002a,b); Vantaggi (2001, 2003) for solving conterintuitive weaknesses of the classical notion of stochastic independence, related to events having either 0 and 1 probability, in Coletti et al. (2020) an analogous approach is used to address correlation between events.

Below we recall the definition of cs-independence given in Coletti and Scozzafava (2002a), and those of positive and negative cs-correlation given in Coletti et al. (2020), where "cs" reads as "coherent setting".

Definition 5. *Let P be a coherent conditional probability defined on an arbitrary family of conditional events $\mathcal{G}$ containing $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$. We say that:*

- *E is* **positively cs-correlated** *with H with respect to P, denoted as $E \perp_{cs}^{+} H$, if one of the following conditions holds:*
  - *$P(E|H) > P(E|H^c)$;*

  - $P(E|H) = P(E|H^c) = 0$, *and there exists a complete agreeing class* $\{P_\alpha\}$ *on* $\boldsymbol{alg}(\{E, H\})$ *that agrees with* $P_{|\mathcal{D}}$ *such that* $o(E|H) < o(E|H^c)$*;*
  - $P(E|H) = P(E|H^c) = 1$, *and there exists a complete agreeing class* $\{P_\alpha\}$ *on* $\boldsymbol{alg}(\{E, H\})$ *that agrees with* $P_{|\mathcal{D}}$ *such that* $o(E^c|H) > o(E^c|H^c)$*;*

- E *is* **negatively cs-correlated** *with* H *with respect to* P*, denoted as* $E \perp_{cs}^{-} H$*, if one of the following conditions holds:*
  - $P(E|H) < P(E|H^c)$*;*
  - $P(E|H) = P(E|H^c) = 0$, *and there exists a complete agreeing class* $\{P_\alpha\}$ *on* $\boldsymbol{alg}(\{E, H\})$ *that agrees with* $P_{|\mathcal{D}}$ *such that* $o(E|H) > o(E|H^c)$*;*
  - $P(E|H) = P(E|H^c) = 1$, *and there exists a complete agreeing class* $\{P_\alpha\}$ *on* $\boldsymbol{alg}(\{E, H\})$ *that agrees with* $P_{|\mathcal{D}}$ *such that* $o(E^c|H) < o(E^c|H^c)$*;*

- E *is* **cs-independent** *of* H *with respect to* P*, denoted as* $E \perp\!\!\!\perp_{cs} H$*, if the following two conditions hold:*
  - $P(E|H) = P(E|H^c)$*;*
  - *there exists a complete agreeing class* $\{P_\alpha\}$ *on* $\boldsymbol{alg}(\{E, H\})$ *that agrees with* $P_{|\mathcal{D}}$ *such that* $o(E|H) = o(E|H^c)$ *and* $o(E^c|H) = o(E^c|H^c)$.

As proved in Coletti and Scozzafava (2002a) and Coletti et al. (2020), the above definitions of stochastic independence and positive/negative correlation in a coherent setting avoid the counterintuitive situations arising from the adoption of the classical definitions. In fact, as shown in the following theorem, in the presence of extreme probability events, these definitions allow the identification of a correlation between events which are logically related, and that, therefore, cannot be stochastically independent.

Theorem 3. *Let* P *be a coherent conditional probability defined on an arbitrary family of conditional events* $\mathcal{G}$ *containing the set* $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$*. Then, the following properties hold:*

(i) if either $E \wedge H = \varnothing$ *or* $E^c \wedge H^c = \varnothing$*, then* $E \perp_{cs}^{-} H$*;*

(ii) if either $E^c \wedge H = \varnothing$ *or* $E \wedge H^c = \varnothing$ *then* $E \perp_{cs}^{+} H$*;*

(iii) if $E \perp\!\!\!\perp_{cs} H$ *then* E *and* H *are logically independent.*

Let $\underline{P}$ be a coherent lower conditional probability defined on an arbitrary set $\mathcal{G}$ of conditional events and $\mathcal{Q}$ the corresponding maximal dominating family. For a non-empty $\mathcal{D} \subseteq \mathcal{G}$, denote by $\mathcal{Q}_{|\mathcal{D}}$ the maximal dominating family related to $\underline{P}_{|\mathcal{D}}$.

Definition 6. *Let* $\underline{P}$ *be a coherent conditional probability defined on an arbitrary family of conditional events* $\mathcal{G}$ *containing* $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$*. Indicating by index* 1 *any element of* $\mathcal{Q}_{|\mathcal{D}}$ *which is i-minimal for* $E|H$ *and by index* 2 *any element of* $\mathcal{Q}_{|\mathcal{D}}$ *which is i-minimal for* $E|H^c$*, we say that:*

- E *is* **positively cs-correlated** *with* H *with respect to* $\underline{P}$ *(denoted as* $E\underline{\perp^+_{cs}}H$*) if* $\exists P^1, P^2 \in \mathcal{Q}_{|\mathcal{D}}$ *such that* $E \perp^+_{cs} H$ *with respect to both and* $\nexists P^{1'}, P^{2'} \in \mathcal{Q}_{|\mathcal{D}}$ *such that either* $E \perp^-_{cs} H$ *with respect to both or* $E \perp\!\!\!\perp_{cs} H$ *with respect to both.*

- E *is* **negatively cs-correlated** *with* H *with respect to* $\underline{P}$ *(denoted as* $E\underline{\perp^-_{cs}}H$*) if* $\exists P^1, P^2 \in \mathcal{Q}_{|\mathcal{D}}$ *such that* $E \perp^-_{cs} H$ *with respect to both and* $\nexists P^{1'}, P^{2'} \in \mathcal{Q}_{|\mathcal{D}}$ *such that either* $E \perp^+_{cs} H$ *with respect to both or* $E \perp\!\!\!\perp_{cs} H$ *with respect to both.*

- E *is* **cs-independent** *of* H *with respect to* $\underline{P}$*, denoted as* $E\underline{\perp\!\!\!\perp_{cs}}H$*, if* $\exists P^1, P^2 \in \mathcal{Q}_{|\mathcal{D}}$ *such that* $E \perp\!\!\!\perp_{cs} H$ *with respect to both and* $\nexists P^{1'}, P^{2'} \in \mathcal{Q}_{|\mathcal{D}}$ *such that either* $E \perp^+_{cs} H$ *with respect to both or* $E \perp^-_{cs} H$ *with respect to both.*

In Coletti and Scozzafava (2002a) a definition of cs-independence for a coherent conditional lower probability $\underline{P}$ has been already introduced. Differently from the present definition, it requires the cs-independence for all the elements of a finite dominating class $\mathcal{P}$. However, the quoted definition is too demanding, since it involves too many conditional events. The strength of this definition is testified by the fact that it is satisfied only if $\underline{P}$, restricted to $\{E, H, E^*|H^*, H^*|E^*\}$, is a coherent conditional probability (Coletti and Scozzafava, 2002a).

We point out that the two elements of pairs P^1, P^2 and $P^{1'}, P^{2'}$ are not necessarily distinct, in the sense that the same coherent conditional probability on $\mathcal{D}$ can be i-minimal for both $E|H$ and $E|H^c$, thus it can happen $P^1 = P^2$ and $P^{1'} = P^{2'}$.

From Definition 6 and Theorems 3 the following Corollary 1 immediately follows.

Corollary 1. *Let* $\underline{P}$ *be a coherent lower conditional probability defined on an arbitrary family of conditional events* $\mathcal{G}$ *containing the set* $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$*. Then, the following properties hold:*

(i) if either $E \wedge H = \varnothing$ *or* $E^c \wedge H^c = \varnothing$*, then* $E\underline{\perp^-_{cs}}H$*;*

(ii) if either $E^c \wedge H = \varnothing$ *or* $E \wedge H^c = \varnothing$ *then* $E\underline{\perp^+_{cs}}H$*;*

(iii) if $E\underline{\perp\!\!\!\perp_{cs}}H$ *then* E *and* H *are logically independent.*

Thanks to the above Corollary, in the following we need to consider only pairs of logically independent events.

In addressing a characterization of relations $\underline{\perp\!\!\!\perp_{cs}}$, $\underline{\perp^+_{cs}}$, $\underline{\perp^-_{cs}}$, we start considering the case in which $\underline{P}(E|H)$ and $\underline{P}(E|H^c)$ do not take extreme values 0 and 1.

Theorem 4. *Let* E, H *be logically independent events and let* $\underline{P}$ *be a coherent lower conditional probability defined on an arbitrary set of conditional events* $\mathcal{G}$ *containing the set* $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$ *and such that* $\underline{P}(E|H) = \alpha$ *and* $\underline{P}(E|H^c) = \beta$*, with* α *and* β *in* $]0, 1[$*. Then the following statements hold:*

(i) $E\underline{\perp\!\!\!\perp_{cs}}H$ *if and only if* $\alpha = \beta$*, and there is* $\delta \in [0, 1]$ *such that*

$$\max\{\underline{P}(H|E), \underline{P}(H|E^c)\} \leq \delta \quad \textit{and} \quad \max\{\underline{P}(H^c|E), \underline{P}(H^c|E^c)\} \leq 1 - \delta;$$

(ii) $E\underline{\perp}_{cs}^{+}H$ *if and only if* $\alpha > \beta$, *and there exist* $\beta' \in [\beta, \alpha[$ *and* $\delta \in [0,1]$ *such that*

$$\begin{aligned} \underline{P}(H|E) &\leq \tfrac{\alpha\delta}{\alpha\delta+\beta'(1-\delta)}, & \underline{P}(H^c|E) &\leq \tfrac{\beta'(1-\delta)}{\alpha\delta+\beta'(1-\delta)}, \\ \underline{P}(H|E^c) &\leq \tfrac{(1-\alpha)\delta}{(1-\alpha)\delta+(1-\beta')(1-\delta)}, & \underline{P}(H^c|E^c) &\leq \tfrac{(1-\beta')(1-\delta)}{(1-\alpha)\delta+(1-\beta')(1-\delta)}; \end{aligned}$$

(iii) $E\underline{\perp}_{cs}^{-}H$ *if and only if* $\alpha < \beta$, *and there exist* $\alpha' \in [\alpha, \beta[$ *and* $\delta \in [0,1]$ *such that*

$$\begin{aligned} \underline{P}(H|E) &\leq \tfrac{\alpha'\delta}{\alpha'\delta+\beta(1-\delta)}, & \underline{P}(H^c|E) &\leq \tfrac{\beta(1-\delta)}{\alpha'\delta+\beta(1-\delta)}, \\ \underline{P}(H|E^c) &\leq \tfrac{(1-\alpha')\delta}{(1-\alpha')\delta+(1-\beta)(1-\delta)}, & \underline{P}(H^c|E^c) &\leq \tfrac{(1-\beta)(1-\delta)}{(1-\alpha')\delta+(1-\beta)(1-\delta)}. \end{aligned}$$

Proof. Statement *(i)* has been proved in Coletti et al. (In press). We prove only statement *(ii)* since the proof of statement *(iii)* is similar.

Let P^1, P^2 be coherent conditional probabilities on $\mathcal{D}$ which are i-minimal for $E|H$ and $E|H^c$, respectively. Then we have $P^1(E|H) = \alpha$, $P^1(E|H^c) \geq \beta$, $P^2(E|H) \geq \alpha$, $P^2(E|H^c) = \beta$. Let $\mathbf{at}(\{E,H\}) = \{C_1, C_2, C_3, C_4\}$ with $C_1 = E \wedge H, C_2 = E \wedge H^c, C_3 = E^c \wedge H, C_4 = E^c \wedge H^c$.

If $\alpha = \beta$ or $\alpha < \beta$, then we cannot find a pair P^1, P^2 such that $E \perp_{cs}^{+} H$ with respect to both, so, $\alpha > \beta$ is necessary to have $E\underline{\perp}_{cs}^{+}H$. If $\alpha > \beta$, then we always have $1 \geq P^2(E|H) > P^2(E|H^c) > 0$ for which $E \underline{\perp}_{cs}^{+} H$, therefore, we cannot have a pair P^1, P^2 such that $E \perp\!\!\!\perp_{cs} H$ with respect to both, nor a pair P^1, P^2 such that $E \perp_{cs}^{-} H$ with respect to both. Thus, it remains to prove the existence of a pair P^1, P^2 such that $E \perp_{cs}^{+} H$ with respect to both. Since P^2 always satisfies such condition, we only need to consider P^1 that satisfies $P^1(E|H) > P^1(E|H^c)$. Let δ be an arbitrary number in $]0,1[$. Below we report all the possible complete agreeing classes $\{P^1_\alpha\}$ on $\mathbf{alg}(\{E,H\})$ that agree with $P^1(E|H) = \alpha$ and $P^1(E|H^c) = \beta' \in [\beta, \alpha[$:

(A)

	C_1	C_2	C_3	C_4
P^1_0	α	0	$1-\alpha$	0
P^1_1	$\bullet$	β'	$\bullet$	$1-\beta'$

$P^1(H|E) = 1$
$P^1(H|E^c) = 1$

(B)

	C_1	C_2	C_3	C_4
P^1_0	0	β'	0	$1-\beta'$
P^1_1	α	$\bullet$	$1-\alpha$	$\bullet$

$P^1(H|E) = 0$
$P^1(H|E^c) = 0$

(C)

	C_1	C_2	C_3	C_4
P^1_0	$\alpha\delta$	$\beta'(1-\delta)$	$(1-\alpha)\delta$	$(1-\beta')(1-\delta)$

$P^1(H|E) = \frac{\alpha\delta}{\alpha\delta+\beta'(1-\delta)}$
$P^1(H|E^c) = \frac{(1-\alpha)\delta}{(1-\alpha)\delta+(1-\beta')(1-\delta)}$

Therefore, since the last two expressions of $P^1(H|E), P^1(H|E^c)$ reduce to 0 and 1 when $\delta = 0$ and $\delta = 1$, respectively, the claim follows. □

Now we consider the case in which $\underline{P}(E|H), \underline{P}(E|H^c)$ can take the extreme value 1. We start recalling the following theorem proved in Coletti et al. (In press), that considers the case $\underline{P}(E|H) = \underline{P}(E|H^c) = 1$.

Theorem 5. *Let* E, H *be logically independent events and let* $\underline{P}$ *be a coherent lower conditional probability defined on an arbitrary set of conditional events* $\mathcal{G}$ *containing the set* $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$ *such that* $\underline{P}(E|H) = \underline{P}(E|H^c) = 1$. *Then the following statements hold:*

(i) $E \underline{\perp\!\!\!\perp}_{cs} H$ *if and only if one of the following conditions holds:*

(a) $0 < \underline{P}(H|E) < 1$, $0 < \underline{P}(H^c|E) < 1$, $0 < \underline{P}(H|E^c) < 1$ *and* $0 < \underline{P}(H^c|E^c) < 1$*;*

(b) $\underline{P}(H|E) = \underline{P}(H|E^c) = 1$*;*

(c) $\underline{P}(H^c|E) = \underline{P}(H^c|E^c) = 1$*;*

(ii) $E \underline{\perp}^+_{cs} H$ *if and only if* $0 < \underline{P}(H|E) \leq 1$ *and* $0 < \underline{P}(H^c|E^c) \leq 1$ *and*

$$\max\{\underline{P}(H|E), \underline{P}(H^c|E^c)\} = 1;$$

(iii) $E \underline{\perp}^-_{cs} H$ *if and only if* $0 < \underline{P}(H^c|E) \leq 1$ *and* $0 < \underline{P}(H|E^c) \leq 1$ *and*

$$\max\{\underline{P}(H^c|E), \underline{P}(H|E^c)\} = 1.$$

Next we consider the case in which one between $\underline{P}(E|H)$ and $\underline{P}(E|H^c)$ takes the extreme value 1 and the other is in the open interval $]0, 1[$.

Theorem 6. *Let* E, H *be logically independent events and let* $\underline{P}$ *be a coherent lower conditional probability defined on an arbitrary set of conditional events* $\mathcal{G}$ *containing* $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$ *and such that* $\underline{P}(E|H) = \alpha$ *and* $\underline{P}(E|H^c) = \beta$*, with* $\alpha, \beta \in]0, 1]$*,* $\max\{\alpha, \beta\} = 1$ *and* $\alpha \neq \beta$*. Then the following statements hold:*

(i) $E \underline{\perp}^+_{cs} H$ *if and only if* $1 = \alpha > \beta > 0$ *and*

$$\max\{\underline{P}(H|E), \underline{P}(H^c|E^c)\} = 1;$$

(ii) $E \underline{\perp}^-_{cs} H$ *if and only if* $0 < \alpha < \beta = 1$ *and*

$$\max\{\underline{P}(H^c|E), \underline{P}(H|E^c)\} = 1.$$

Proof. We prove only statement *(i)* since the proof of statement *(ii)* is similar.

Let P^1, P^2 be coherent conditional probabilities on $\mathcal{D}$ which are i-minimal for $E|H$ and $E|H^c$, respectively. Then we have $P^1(E|H) = \alpha$, $P^1(E|H^c) \geq \beta$, $P^2(E|H) \geq \alpha$, $P^2(E|H^c) = \beta$. Let $\mathbf{at}(\{E, H\}) = \{C_1, C_2, C_3, C_4\}$ with $C_1 = E \wedge H, C_2 = E \wedge H^c, C_3 = E^c \wedge H, C_4 = E^c \wedge H^c$.

Since we assume $\alpha \neq \beta$, if $\alpha < \beta$, then we cannot find a pair P^1, P^2 such that $E \perp^+_{cs} H$ with respect to both, so, $\alpha > \beta$ is necessary to have $E \underline{\perp}^+_{cs} H$. If $1 = \alpha > \beta > 0$, then we always have $1 = P^2(E|H) > P^2(E|H^c) > 0$ for which $E \perp^+_{cs} H$, therefore, we cannot have a pair P^1, P^2 such that $E \perp\!\!\!\perp_{cs} H$ with respect to both, nor a pair P^1, P^2 such that $E \perp^-_{cs} H$ with respect to both. Thus, it remains to prove the existence of a pair P^1, P^2 such that $E \perp^+_{cs} H$ with respect to both. Since P^2 always satisfies such condition, we only need to consider P^1 that satisfies $P^1(E|H) > P^1(E|H^c)$ or $P^1(E|H) = P^1(E|H^c) = 1$ and $o(E^c|H) > o(E^c|H^c)$. Let δ be an arbitrary number in $]0, 1[$. Below we report all the possible complete agreeing classes $\{P^1_\alpha\}$ on $\mathbf{alg}(\{E, H\})$ that agree with $P^1(E|H) = 1$ and $P^1(E|H^c) = \beta' \in [\beta, 1[$:

(A)

	C_1	C_2	C_3	C_4
P_0^1	1	0	0	0
P_1^1	$\bullet$	0	1	0
P_2^1	$\bullet$	β'	$\bullet$	$1-\beta'$

$P^1(H|E) = 1$
$P^1(H|E^c) = 1$

(B)

	C_1	C_2	C_3	C_4
P_0^1	1	0	0	0
P_1^1	$\bullet$	β'	0	$1-\beta'$
P_2^1	$\bullet$	$\bullet$	1	$\bullet$

$P^1(H|E) = 1$
$P^1(H|E^c) = 0$

(C)

	C_1	C_2	C_3	C_4
P_0^1	1	0	0	0
P_1^1	$\bullet$	$\beta'(1-\delta)$	δ	$(1-\beta')(1-\delta)$

$P^1(H|E) = 1$
$P^1(H|E^c) = \frac{\delta}{\delta+(1-\beta')(1-\delta)}$

(D)

	C_1	C_2	C_3	C_4
P_0^1	0	β'	0	$1-\beta'$
P_1^1	1	$\bullet$	0	$\bullet$
P_2^1	$\bullet$	$\bullet$	1	$\bullet$

$P^1(H|E) = 0$
$P^1(H|E^c) = 0$

(E)

	C_1	C_2	C_3	C_4
P_0^1	δ	$\beta'(1-\delta)$	0	$(1-\beta')(1-\delta)$
P_1^1	$\bullet$	$\bullet$	1	$\bullet$

$P^1(H|E) = \frac{\delta}{\delta+\beta'(1-\delta)}$
$P^1(H|E^c) = 0$

For every value of $\beta' \in [\beta, 1[$ we notice that the expressions $\frac{\delta}{\delta+(1-\beta')(1-\delta)}$ and $\frac{\delta}{\delta+\beta'(1-\delta)}$ can take any value in $]0,1[$ by varying $\delta \in]0,1[$. In particular, they reduce to 0 and 1 by taking $\delta = 0$ and $\delta = 1$.

Below we report all the possible complete agreeing classes $\{P_\alpha^1\}$ on $\mathbf{alg}(\{E, H\})$ that agree with $P^1(E|H) = P^1(E|H^c) = 1$ such that $o(E^c|H) > o(E^c|H^c)$:

(F)

	C_1	C_2	C_3	C_4
P_0^1	δ	$1-\delta$	0	0
P_1^1	$\bullet$	$\bullet$	0	1
P_2^1	$\bullet$	$\bullet$	1	$\bullet$

$o(E^c|H) = 2 > 1 = o(E^c|H^c)$
$P^1(H|E) = \delta$
$P^1(H|E^c) = 0$

(G)

	C_1	C_2	C_3	C_4
P_0^1	1	0	0	0
P_1^1	$\bullet$	1	0	0
P_2^1	$\bullet$	$\bullet$	0	1
P_3^1	$\bullet$	$\bullet$	1	$\bullet$

$o(E^c|H) = 3 > 1 = o(E^c|H^c)$
$P^1(H|E) = 1$
$P^1(H|E^c) = 0$

(H)

	C_1	C_2	C_3	C_4
P_0^1	1	0	0	0
P_1^1	$\bullet$	1	0	0
P_2^1	$\bullet$	$\bullet$	δ	$1-\delta$

$o(E^c|H) = 2 > 1 = o(E^c|H^c)$
$P^1(H|E) = 1$
$P^1(H|E^c) = \delta$

Hence, as soon as we can find a P^1 such that $P^1(H|E) = 1$ we do not have any constraint on the values of $P^1(H|E^c)$ and, analogously, as soon as we can find a P^1 such that $P^1(H^c|E^c) = 1$ we do not have any constraint on the values of $P^1(H|E)$. Therefore, $E \underline{\perp}_{cs}^{+} H$ if and only if $1 = \alpha > \beta > 0$ and $\max\{\underline{P}(H|E), \underline{P}(H^c|E^c)\} = 1$. $\square$

4 Conclusions

The classic definitions of stochastic independence and positive and negative correlation for events fail when one considers two events of extreme probability. In fact, in this circumstance, events result to be independent even in the case they are logically dependent.

To remove this inconvenience, in Coletti and Scozzafava (2002b) and Coletti and Scozzafava (2002a) a stronger definition of independence (named cs-independence) was introduced. It adds to the most stringent classical definition (i.e., $P(E|H) = P(E|H^c)$) a condition on the zero-layers of $E|H, E|H^c$ and $E^c|H, E^c|H^c$, which comes into operation when $P(E|H) = P(E|H^c)$ are equal to 0 or 1. Following the same idea, in Coletti et al. (2020) reinforced definitions of positive and negative correlation have been introduced. To remove difficulties due to the management of zero-layers, in the above articles a complete characterization of independence and correlation only involving the values of $P(E^*|H^*), P(H^*|E^*), P(H)$ has been proposed. Some typos related to the characterization of positive and negative correlation present in Coletti et al. (2020) have been corrected in Coletti et al. (In press). Similar results are obtained in Coletti and Vantaggi (2006) for independence in the framework of conditional possibility.

In this paper we present the extension of cs-independence and cs-correlation for coherent lower conditional probability. Actually, a tentative extension of cs-independence to lower conditional probability is present in the literature (see Coletti and Scozzafava (2002a), but it results to be too strong, so that it implies that two events E and H are cs-independent for a coherent lower conditional probability $\underline{P}$ if and only if the restriction of $\underline{P}$ to $\{E, H, E^*|H^*, H^*|E^*\}$ is a coherent conditional probability. The present definitions of cs-independence and cs-correlation are based on local conditions. In details, they require the existence of a pair of i-minimal coherent conditional probabilities for $E|H$ and $E|H^c$, respectively, that unanimously agree on cs-independence or cs-correlation of E and H, and the non-existence of a pair of i-minimal coherent conditional probabilities for $E|H$ and $E|H^c$, respectively, that unanimously agree on a different statement of cs-correlation or cs-independence.

We also gave a characterization of the above relations in terms of lower conditional probability restricted to $\mathcal{D} = \{E^*|H^*, H^*|E^*\}$. Due to the lack of space we presented only the case where $\underline{P}(E|H) = \alpha$ and $\underline{P}(E|H) = \beta$ with $\alpha = \beta \in]0,1[$, that with $1 = \alpha > \beta > 0$, and that with $0 < \alpha < \beta = 1$. Moreover we reported the case where $\underline{P}(E|H) = \underline{P}(E|H^c) = 1$, proved in Coletti et al. (In press). We did not present the cases where $\underline{P}(E|H), \underline{P}(E|H^c)$ can take 0, since proofs do not fit the length restrictions.

References

G. Coletti and R. Scozzafava. Stochastic Independence in a Coherent Setting. *Annals of Mathematics and Artificial Intelligence*, 35(1):151–176, 2002a.

G. Coletti and R. Scozzafava. *Technologies for Constructing Intelligent Systems 2*, volume 90 of *Studies in Fuzziness and Soft Computing*, chapter Stochastic Independence for Upper and Lower Probabilities in a Coherent Setting, pages 17–30. Physica, Heidelberg, 2002b.

G. Coletti and R. Scozzafava. *Probabilistic Logic in a Coherent Setting*, volume 15 of *Trends in Logic*. Kluwer Academic Publisher, Dordrecht/Boston/London, 2002c.

G. Coletti and B. Vantaggi. Possibility theory: Conditional independence. *Fuzzy Sets and Systems*, 157(11):1491–1513, 2006.

G. Coletti, L. van der Gaag, D. Petturiti, and B. Vantaggi. Detecting correlation between extreme probability events. *International Journal of General Systems*, 49(1):64–87, 2020.

G. Coletti, S. Latini, and D. Petturiti. A discussion about independence and correlation in the framework of coherent lower conditional probability. In D. Ciucci, I. Couso, J. Medina, D. Slezak, D. Petturiti, B. Bouchon-Meunier, and R. Yager, editors, *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, CCIS. Springer International Publishing, In press.

I. Couso, S. Moral, and P. Walley. Examples of independence for imprecise probabilities. In *Intternationa Symposium on Imprecise Probabilities and their applications (ISIPTA '99), Ghent, Belgium*, pages 121—-130, 1999.

L. Dubins. Finitely Additive Conditional Probabilities, Conglomerability and Disintegrations. *The Annals of Probability*, 3(1):89–99, 1975.

B. Vantaggi. Conditional independence in a coherent finite setting. *Annals of Mathematics and Artificial Intelligence*, 32(1):287–313, 2001.

B. Vantaggi. Conditional independence structures and graphical models. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(05):545–571, 2003.

P. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44(3):366–383, 2007.

A look to independence under T-conditional possibility

Giulianella Coletti[1], Davide Petturiti[2], and Barbara Vantaggi[3]

[1]Dept. Mathematics and Computer Science, University of Perugia
giulianella.coletti@unipg.it
[2]Dept. Economics, University of Perugia
davide.petturiti@unipg.it
[3]Dept. MEMOTEF, "La Sapienza" University of Rome
barbara.vantaggi@uniroma1.it

Abstract

The notion of conditioning is still debated for non-additive uncertainty measures. Referring to the possibilistic framework, we consider different notions of conditioning, focusing on the axiomatic definition of T-conditional possibility that can accommodate Dubois and Prade's conditioning rule. A notion strictly linked to that of conditioning is independence, for which we provide a comparison with respect to the different conditioning rules. In particular, we introduce conditional independence for variables under T-conditional possibility, with T a continuous t-norm, by taking as a significant particular case T_{DP}-conditional possibility (obtained through Dubois and Prade's minimum specificity principle).

1 Introduction

The notion of conditioning is a problem of long-standing interest and it involves different uncertainty measures. In this paper we focus on possibility theory where various definitions of conditional possibilities have been introduced: by analogy with the Kolmogorovian probabilistic framework or by using some criterion as the minimum specificity principle (see e.g. Benferhat et al. (2011); de Cooman (1997a); Dubois and Prade (1988)). All the provided definitions have in common the fact that the conditional measure is obtained as a derived concept from the "unconditional" one.

In Bouchon-Meunier et al. (2002) a general notion of T-conditional possibility has been introduced as a primitive concept: the conditional possibility is directly defined as a function on a set of conditional events which satisfies a suitable set of axioms and it is not induced just by a single unconditional possibility (as solution of an equation involving joint and marginal possibilities). Characterizations of T-conditional possibilities have been provided in Coletti and Vantaggi (2006); Ferracuti and Vantaggi (2006); Coletti and Vantaggi (2009) and a comparison with the conditioning notion obtained through the minimum specificity principle, called here T_{DP}-conditioning, introduced in Dubois and Prade (1988), is present in Coletti et al. (2013).

For each of these notions of conditional possibilities one can consider an ensuing notion of independence, which is reinforced with respect to the classical one.

The main motivations of the reinforcement of the independence notions is to capture the following natural implication: the independence under an uncertainty measure (and so, in particular, under a possibility) must imply logical independence. In other words if an event is "logically" related to another one, the two events cannot be independent under any uncertainty measure. This implication, even though very intuitive, can fail when we adopt the classical definitions of independence. We note that, taking into account logical constraints is interesting not only from a theoretical point of view, but also in applications.

This work aims to compare the effect of different conditional possibility definitions on the reinforced independence notion, focusing on T-conditional possibility and T_{DP}-conditional possibility, with T any continuous t-norm.

In particular, we analyze the conditional independence notion studied in Coletti and Vantaggi (2006, 2009) under T-conditional possibility and T_{DP}-conditional possibility. A comparison is provided for the t-norm of the minimum and for strict t-norms.

2 Conditioning in possibility theory

In what follows, $\mathcal{B}\times\mathcal{H}$ denotes a set of conditional events with $\mathcal{B}$ a Boolean algebra and $\mathcal{H}$ an additive set (i.e., closed with respect to finite logical sums) such that $\mathcal{H} \subseteq \mathcal{B}^0 = \mathcal{B}\setminus\{\emptyset\}$.

Moreover, given a finite set $\mathcal{G} = \{E_i|H_i\}_{i=1,\dots,n}$, let $\langle\{E_i, H_i\}_{i=1,\dots,n}\rangle$ be the algebra spanned by the events $\{E_i, H_i\}_{i=1,\dots,n}$ and $\mathcal{C}_{\langle\{E_i,H_i\}_{i=1,\dots,n}\rangle}$ the relevant set of atoms.

Definition 1. *Let T be any t-norm. A function $\Pi : \mathcal{B}\times\mathcal{H} \to [0,1]$ is a* **T-conditional possibility** *if it satisfies the following properties:*

(CP1) $\Pi(E|H) = \Pi(E \wedge H|H)$*, for every $E \in \mathcal{B}$ and $H \in \mathcal{H}$;*

(CP2) $\Pi(\cdot|H)$ *is a finitely maxitive possibility on $\mathcal{B}$, for any $H \in \mathcal{H}$;*

(CP3) $\Pi(E \wedge F|H) = T(\Pi(E|H), \Pi(F|E \wedge H))$*, for any $H, E \wedge H \in \mathcal{H}$ and $E, F \in \mathcal{B}$.*

An equivalent set of axioms is obtained by replacing *(CP1)* with the following *(CP1')*:

(CP1') $\Pi(H|H) = 1$, for every $H \in \mathcal{H}$.

Let us stress that condition *(CP2)* requires that, for every $H \in \mathcal{H}$, $\Pi(\cdot|H)$ is a normalized finitely maxitive function (Shilkret, 1971) defined on $\mathcal{B}$.

A T-conditional possibility is *full* if it is defined on $\mathcal{B}\times\mathcal{B}^0$. In the following T is assumed to be continuous, since for not continuous t-norms the extendibility as a full T-conditional possibility is not assured (see Example 1 in Coletti and Vantaggi (2009)).

A full T-conditional possibility $\Pi(\cdot|\cdot)$ on $\mathcal{B}$ is not necessarily "represented" by means of a single unconditional possibility, even when $\mathcal{B}$ is finite (see Coletti and Vantaggi (2009)): in this case, there is a (unique) class of possibility distributions $\mathcal{P} = \{\Pi_0, \ldots, \Pi_k\}$ defined on $\mathcal{C}_{\mathcal{B}}$, called *$T$-nested class* agreeing with $\Pi(\cdot|\cdot)$, that is such that, for $\alpha = 1, \ldots, k$:

1. $\Pi_{\alpha-1}(C) \leq \Pi_\alpha(C)$ if $C \in \mathcal{C}_\alpha$,
2. $\Pi_\alpha(C) = 0$ for all the atoms $C \in \mathcal{C}_0 \setminus \mathcal{C}_\alpha$,
3. for any $C \in \mathcal{C}_0$ there exists a (unique) $\alpha_C \in \{0, \ldots, k\}$ such that $\Pi_{\alpha_C}(C) = 1$,
4. for any $C_1, C_2 \in \mathcal{C}_\alpha$, $\Pi_{\alpha-1}(C_1) < \Pi_{\alpha-1}(C_2) \Longrightarrow \Pi_\alpha(C_1) < \Pi_\alpha(C_2)$,
5. for any $C \in \mathcal{C}_\alpha$, $\Pi_{\alpha-1}(C) = T(\Pi_\alpha(C), \Pi_{\alpha-1}(H_0^\alpha))$.

where $\mathcal{C}_0 = \mathcal{C}_{\mathcal{B}}$, $\mathcal{C}_\alpha = \{C \in \mathcal{C}_{\alpha-1} : \Pi_{\alpha-1}(C) < 1\}$ and $H_0^\alpha = \bigvee_{C \in \mathcal{C}_\alpha} C$.

The class $\mathcal{P}$ induces a "layered" partition of the algebra $\mathcal{B}$ that allows to define a rank among the events in $\mathcal{B}$, that can be extended to all the conditional events in $\mathcal{B} \times \mathcal{B}^0$.

Note that the above definition of agreeing class differs from that given in Ferracuti and Vantaggi (2006) for T-conditional possibility with T a strict t-norm.

This highlights an important difference with other approaches to conditioning, where the conditional possibility $\Pi(E|H)$ is *defined*, starting from an unconditional possibility $\Pi(\cdot)$, as a solution of the equation in x

$$\Pi(E \wedge H) = T(x, \Pi(H)). \tag{1}$$

Continuity of T assures only the solvability of (1), while to reach the uniqueness of the solution a further constraint must be imposed. At this aim, the Dubois and Prade's *minimum specificity principle* (Dubois and Prade, 1988) consists in always selecting the greatest solution of equation (1), by means of the residuum $\to_T$ of a continuous t-norm, defined as

$$x \to_T y = \sup\{z \in [0,1] : T(x,z) \leq y\}.$$

By referring to this notion of conditioning (see, e.g., Dubois and Prade (1988)) we deal with T_{DP}-conditional possibility.

Definition 2. *Let T be a continuous t-norm. A function $\Pi : \mathcal{B} \times \mathcal{H} \to [0,1]$ is a T_{DP}-conditional possibility if it satisfies the following conditions:*

(DP1) $\Pi(E|H) = \Pi(E \wedge H|H)$ *for every* $E \in \mathcal{B}$ *and* $H \in \mathcal{H}$*;*

(DP2) $\Pi(\cdot|H)$ *is a finitely maxitive possibility for every* $H \in \mathcal{H}$*;*

(DP3) for every $E|H \in \mathcal{B} \times \mathcal{H}$ it holds (with $H_0^0 = \bigvee_{H \in \mathcal{H}} H \in \mathcal{H}$)

$$\Pi(E|H) = \begin{cases} \Pi(H|H_0^0) \to_T \Pi(E \wedge H|H_0^0) & \text{if } E \wedge H \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

Next proposition shows that T_{DP}-conditional possibilities are particular T-conditional possibilities.

Proposition 1. *Let T be a continuous t-norm. If Π on $\mathcal{B} \times \mathcal{H}$ is a T_{DP}-conditional possibility, then Π is a T-conditional possibility.*

The above definitions deeply rely on the Boolean structure of the domain of the function Π, thus in order to remove any restriction on the domain we go back to the concept of *coherence*, originally introduced by de Finetti for (finitely additive) probabilities (de Finetti, 1949).

Definition 3. *Let T be a t-norm. Given a set $\mathcal{G} = \{E_i|H_i\}_{i=1,\dots,n}$ of conditional events, an assessment $\Pi : \mathcal{G} \to [0,1]$ is a* **coherent T-conditional [T_{DP}-conditional] possibility** *if there exits a full T-conditional [T_{DP}-conditional] possibility Π' on $\mathcal{B} = \langle\{E_i, H_i\}_{i=1,\dots,n}\rangle$ extending Π.*

As shown in Coletti et al. (2013), Proposition 1 implies that any coherent T_{DP}-conditional possibility is a coherent T-conditional possibility, but the converse does not hold.

Let us stress that not all the definitions of conditioning present in the literature are particular T-conditional possibilities: this is the case of Zadeh's conditioning rule (see Coletti and Vantaggi (2006)). Actually, T-conditional possibilities according to Definition 1 are consistent with the definition given in de Cooman (1997a) which deals with the problem of conditioning by following the classic Kolmogorovian line, defining for this purpose the concept of (Π, T)-almost everywhere equality. Nevertheless, the class of conditional measures consistent with the definition given in de Cooman (1997a) contains also functions not consistent with our Definition 1, in the sense that, for some conditioning event H we can have $\Pi(H|H) \neq 1$.

Now we study T_{DP}-conditional possibility in the case the Boolean algebra $\mathcal{B}$ is finite. In the next proposition we show that every T_{DP}-conditional possibility on $\mathcal{B} \times \mathcal{H}$ can be extended (not necessarily in a unique way) to a full T_{DP}-conditional possibility on $\mathcal{B}$ (i.e., a T_{DP}-conditional possibility on $\mathcal{B} \times \mathcal{B}^0$).

Proposition 2. *Let T be a continuous t-norm, and $\mathcal{B}$ a finite algebra. If $\Pi : \mathcal{B} \times \mathcal{H} \to [0,1]$ is a T_{DP}-conditional possibility, then there exists a full T_{DP}-conditional possibility $\Pi' : \mathcal{B} \times \mathcal{B}^0 \to [0,1]$ such that $\Pi'_{|\mathcal{B}\times\mathcal{H}} = \Pi$.*

Since a full T_{DP}-conditional possibility is a particular full T-conditional possibility, it can be "represented" by means of a unique T-nested class $\mathcal{P} = \{\Pi_0, \dots, \Pi_k\}$ agreeing with it (see Coletti and Vantaggi (2009)).

Remark 1. *By referring to the T-nested class of a full T_{DP}-conditional possibility, note that given $\Pi_0(\cdot) = \Pi(\cdot|\Omega)$, if Π_0 takes k distinct values $1 > \pi_1 > \pi_2 > \ldots > \pi_k \geq 0$, then for $\alpha = 1, \dots, k$, the distribution of Π_α is obtained assigning $\Pi_\alpha(C_r) = 0$ to all those atoms $C_r \notin \mathcal{C}_\alpha$ and $\Pi_\alpha(C_r) = \pi_\alpha \to_T \Pi_{\alpha-1}(C_r)$ for all the atoms $C_r \in \mathcal{C}_\alpha$. Therefore the specificity of T_{DP}-conditional possibilities can be captured directly through the particular structure of the corresponding T-nested classes.*

In the next proposition we show that every full T_{DP}-conditional possibility on $\mathcal{B}$ can be extended as a full T_{DP}-conditional possibility on every finite superalgebra $\mathcal{B}' \supseteq \mathcal{B}$.

Proposition 3. *Let T be a continuous t-norm, $\mathcal{B}$ a finite algebra and $\mathcal{B}' \supseteq \mathcal{B}$ a finite superalgebra. If $\Pi : \mathcal{B} \times \mathcal{B}^0 \to [0,1]$ is a full T_{DP}-conditional possibility, then there exists a full T_{DP}-conditional possibility $\Pi' : \mathcal{B}' \times \mathcal{B}'^0 \to [0,1]$ such that $\Pi'_{|\mathcal{B}\times\mathcal{B}^0} = \Pi$.*

In Coletti and Vantaggi (2009) the coherence of a T-conditional possibility assessment Π on a finite $\mathcal{G}$ has been characterized also in terms of a proper sequence of compatible systems $\mathcal{S}_0^\Pi, \ldots, \mathcal{S}_k^\Pi$, whose solutions are the possibility distributions related to a T-nested class of possibilities $\mathcal{P} = \{\Pi_0, \ldots, \Pi_k\}$.

A similar characterization can be proved for coherent T_{DP}-conditional possibilities.

Theorem 1. *Let T be a continuous t-norm and $\mathcal{G} = \{E_i|H_i\}_{i=1,\ldots,n}$. For a function $\Pi : \mathcal{G} \to [0,1]$, the following statements are equivalent:*

(a) Π is a coherent T_{DP}-conditional possibility on $\mathcal{G}$;

(b) for any $E_i|H_i \in \mathcal{G}$ such that $E_i \wedge H_i = \emptyset$, it is $\Pi(E_i|H_i) = 0$, and the following system with unknowns $x_r \geq 0$ for $C_r \in \mathcal{C}_0 = \mathcal{C}_{\langle\{E_i,H_i\}_{i=1,\ldots,n}\rangle}$, is compatible

$$\mathcal{S}_T^{DP} = \begin{cases} \max\limits_{C_r \subseteq H_i} x_r \to_T \max\limits_{C_r \subseteq E_i \wedge H_i} x_r = \Pi(E_i|H_i) & \text{if } E_i \wedge H_i \neq \emptyset \\ \max\limits_{C_r \in \mathcal{C}_0} x_r = 1. \end{cases} \tag{2}$$

3 Possibilistic independence under different t-norms

We extend to T-conditional [T_{DP}-conditional] possibilities a notion of possibilistic independence (introduced in Coletti and Vantaggi (2006); Ferracuti and Vantaggi (2006) for the minimum t-norm and for strict t-norms) able to avoid pathological situations whenever logical constraints are involved. In what follows, E^* stands either for E or E^c.

Let $\mathcal{B}$ be a finite Boolean algebra. In order to introduce conditional independence in a possibilistic framework, for $C \in \mathcal{B}^0$, denote by $\mathcal{B} \wedge C = \{E \wedge C : E \in \mathcal{B}\}$ the Boolean ideal of $\mathcal{B}$ with C for its top element and let $\mathcal{C}_{\mathcal{B}\wedge C}$ be the set of atoms of $\mathcal{B}$ belonging to $\mathcal{B} \wedge C$.

For a continuous t-norm T, we call *T-nested class on $\mathcal{B} \wedge C$* a class of possibility distributions $\mathcal{P} = \{\Pi_0, \ldots, \Pi_k\}$ on $\mathcal{C}_\mathcal{B}$ satisfying the properties 1–5 of a T-nested class for $\mathcal{C}_0 = \mathcal{C}_{\mathcal{B}\wedge C}$ and such that $\Pi_\alpha(C) = 0$, for all $C \notin \mathcal{C}_0$, for $j = 0, \ldots, k$. In particular, taking $C = \Omega$ we get back to the classical definition of T-nested class. It is easily proven (see, e.g., Coletti and Petturiti (2016)) that a T-nested class on $\mathcal{B}\wedge C$ uniquely represents a T-conditional possibility $\Pi'(\cdot|\cdot)$ on $\mathcal{B} \times (\mathcal{B} \wedge C)^0$. In particular, if Π is defined on a non-empty set $\mathcal{G} \subseteq \mathcal{B} \times (\mathcal{B} \wedge C)^0$ and $\Pi'_{|\mathcal{G}} = \Pi$, then the T-nested class $\mathcal{P}$ on $\mathcal{B} \wedge C$ representing Π' is said to *agree* with Π.

Definition 4. *Let $\mathcal{B}$ be a finite Boolean algebra, $C \in \mathcal{B}^0$, and $\mathcal{P} = \{\Pi_0, \ldots, \Pi_k\}$ a T-nested class on $\mathcal{B} \wedge C$. Then, for every event $E \in \mathcal{B}^0$, the C-**significant layer** of E*

*(denoted as $\circ(E)$) related to $\mathcal{P}$ is defined as the minimum index α such that $\Pi_\alpha(E) = 1$. Moreover, define $\circ(\emptyset) = +\infty$. For every $E|H \in \mathcal{B} \times (\mathcal{B} \wedge C)^0$ the C-**significant layer** $\circ(E|H)$ of $E|H$, related to $\mathcal{P}$, is defined as the (non-negative) number $\circ(E|H) = \circ(E \wedge H) - \circ(H)$. In case $C = \Omega$, the Ω-significant layer is simply said* **significant layer**.

The notion of significant layer differs from that of zero-layer given for T-conditional possibility with T a strict t-norm: in Ferracuti and Vantaggi (2006) the zero-layer of an event is defined as the (minimum) layer α where the event as positive possibility Π_α.

Now we are able to introduce a definition of conditional independence.

Definition 5. *Let T be a continuous t-norm, $\mathcal{G}$ a set of conditional events containing $\mathcal{D} = \{A^*|B^* \wedge C, B^*|A^* \wedge C\}$. Given a coherent T-conditional [T_{DP}-conditional] possibility Π on $\mathcal{G}$,* A **is independent of** B **given** C **under** Π, *in symbol $A \perp\!\!\!\perp B \,|\, C\,[\Pi]$, if both the following conditions hold:*

- *(i)* $\Pi(A|B \wedge C) = \Pi(A|B^c \wedge C)$ *and* $\Pi(A^c|B \wedge C) = \Pi(A^c|B^c \wedge C)$;
- *(ii) there exists a T-nested class $\mathcal{P}_\mathcal{D} = \{\Pi_\alpha\}_{\alpha=0}^t$ on $\langle\{A, B, C\}\rangle \wedge C$ agreeing with $\Pi_{|\mathcal{D}}$ such that*

$$\circ(A|B \wedge C) = \circ(A|B^c \wedge C) \text{ and } \circ(A^c|B \wedge C) = \circ(A^c|B^c \wedge C). \tag{3}$$

In the above definition, if we take $C = \Omega$, then we obtain a definition of unconditional independence between events A and B, that we simply write as $A \perp\!\!\!\perp B\,[\Pi]$.

We recall that two events A, B are logically independent if all the events of the form $A^* \wedge B^*$ are possible, furthermore logical independence of two possible events A and B, with respect to a possible event C (i.e. $C \neq \emptyset$), means that all the events of the form $A^* \wedge B^* \wedge C$ must be possible. Note that logical independence with respect to an event implies logical independence.

The next theorem shows the connection between the logical independence and possibilistic independence (according to Definition 5), and this holds for any T-nested class on $\langle\{A, B, C\}\rangle \wedge C$.

Theorem 2. *For any continuous t-norm T and for any coherent T-conditional [T_{DP}-conditional] possibility Π on $\mathcal{G}$ containing $\mathcal{D} = \{A^*|B^* \wedge C, B^*|A^* \wedge C\}$, if $A \perp\!\!\!\perp B \,|\, C\,[\Pi]$, then A and B are logically independent with respect to C.*

Proof. The proof is direct and is based on the fact that if an event of the form $A^* \wedge B^* \wedge C$ is impossible then $\circ(A^*|B^* \wedge C) = +\infty$ for every T-nested class on $\langle\{A, B, C\}\rangle \wedge C$ agreeing with $\Pi_{|\mathcal{D}}$. □

For coherent T-conditional possibility the T-nested class $\mathcal{P}_\mathcal{D}$ on $\langle\{A, B, C\}\rangle \wedge C$ agreeing with $\Pi_{|\mathcal{D}}$ is generally non-unique. Then condition *(ii)* seems to depend on the choice of the class. The next theorem proves that the validity of condition *(ii)* is invariant with respect to the choice of the T-nested class on $\langle\{A, B, C\}\rangle \wedge C$.

Theorem 3. *Let A and B be two logically independent events with respect to an event C, and let Π be a coherent T-conditional possibility (with T any continuous t-norm), defined on $\mathcal{G}$ containing $\mathcal{D} = \{A^*|B^* \wedge C, B^*|A^* \wedge C\}$ such that condition (i) of Definition 5 holds. If there exists a T-nested class on $\langle\{A, B, C\}\rangle \wedge C$ agreeing with $\Pi_{|\mathcal{D}}$ such that equation (3) is satisfied, then equation (3) holds for any T-nested class on $\langle\{A, B, C\}\rangle \wedge C$ agreeing with $\Pi_{|\mathcal{D}}$.*

The following theorem characterizes conditional independence avoiding any direct reference to C-significant layers, in the case of coherent T-conditional possibility with T any continuous t-norm.

Theorem 4. *Let T be any continuous t-norm, and A and B two logically independent events with respect to an event C. If a coherent T-conditional possibility is such that $\Pi(A|B \wedge C) = \Pi(A|B^c \wedge C)$ and $\Pi(A^c|B \wedge C) = \Pi(A^c|B^c \wedge C)$, then $A \perp\!\!\!\perp B \mid C\,[\Pi]$ if and only if one (and only one) of the following conditions holds:*

(a) $\Pi(A|B \wedge C) = \Pi(A^c|B \wedge C) = 1$*;*

(b) $\min\{\Pi(A|B \wedge C), \Pi(A^c|B \wedge C)\} = 0$ *and the coherent extension of* Π *on* $\{B^*|A^* \wedge C\}$ *satisfies one of the following conditions*

- $\Pi(B|A \wedge C) < 1$, $\Pi(B|A^c \wedge C) < 1$*;*
- $\Pi(B^c|A \wedge C) < 1$, $\Pi(B^c|A^c \wedge C) < 1$*;*
- $\Pi(B^*|A^* \wedge C) = 1$*;*

(c) $\Pi(A|B \wedge C) = \alpha \in (0, 1)$ *and the coherent extension of* Π *on* $\{B^*|A^* \wedge C\}$ *satisfies one of the following conditions*

- $\Pi(B|A^* \wedge C) = 0$ *or* $\Pi(B^c|A^* \wedge C) = 0$*;*
- $0 < \Pi(B|A^* \wedge C) < \alpha$ *or* $0 < \Pi(B^c|A^* \wedge C) < \alpha$*;*
- $\Pi(B|A^c \wedge C) = \alpha$ *and* $\alpha \leq \Pi(B|A \wedge C) < 1$ *or* $\Pi(B^c|A^c \wedge C) = \alpha$ *and* $\alpha \leq \Pi(B^c|A \wedge C) < 1$*;*
- $\Pi(B^*|A^* \wedge C) = 1$*;*

(d) $\Pi(A^c|B \wedge C) = \alpha \in (0, 1)$ *and the coherent extension of* Π *on* $\{B^*|A^* \wedge C\}$ *satisfies one of the following conditions*

- $\Pi(B|A^* \wedge C) = 0$ *or* $\Pi(B^c|A^* \wedge C) = 0$*;*
- $0 < \Pi(B|A^* \wedge C) < \alpha$ *or* $0 < \Pi(B^c|A^* \wedge C) < \alpha$*;*
- $\Pi(B|A \wedge C) = \alpha$ *and* $\alpha \leq \Pi(B|A^c \wedge C) < 1$ *or* $\Pi(B^c|A \wedge C) = \alpha$ *and* $\alpha \leq \Pi(B^c|A^c \wedge C) < 1$*;*
- $\Pi(B^*|A^* \wedge C) = 1$.

From previous theorem it is possible to derive also an analogous characterization for coherent T_{DP}-conditional possibilities. This can be done (once $\Pi(A|B^* \wedge C)$ and $\Pi(A^c|B^* \wedge C)$ are fixed) by taking into account only the coherent values for $\Pi(B|A^* \wedge C)$ and $\Pi(B^c|A^* \wedge C)$ with respect to the T_{DP}-conditioning that satisfy condition *(ii)* of Definition 5. In this case the significant layers are implied by Remark 1. Below we consider the case $T = \min$ or a strict t-norm.

Theorem 5. *Let T be the minimum or a strict t-norm, and A and B two logically independent events with respect to an event C. If a coherent T_{DP}-conditional possibility is such that $\Pi(A|B\wedge C) = \Pi(A|B^c\wedge C)$ and $\Pi(A^c|B\wedge C) = \Pi(A^c|B^c\wedge C)$, then $A \perp\!\!\!\perp B \mid C\,[\Pi]$ if and only if one (and only one) of the following conditions holds:*

- *(a)* $\Pi(A|B \wedge C) = \Pi(A^c|B \wedge C) = 1$;
- *(b)* $\min\{\Pi(A|B\wedge C), \Pi(A^c|B\wedge C)\} = 0$ *and the coherent extension of* Π *on* $\{B^*|A^*\wedge C\}$ *is such that* $\Pi(B^*|A^* \wedge C) = 1$;
- *(c)* $\min\{\Pi(A|B \wedge C), \Pi(A^c|B \wedge C)\} = \alpha \in (0, 1)$ *and the coherent extension of* Π *on* $\{B^*|A^* \wedge C\}$ *is such that:*
  - *(i)* $\Pi(B^*|A^* \wedge C) = 1$, *if* $T = \min$;
  - *(ii)* $\Pi(B|A \wedge C) = \Pi(B|A^c \wedge C) > 0$ *and* $\Pi(B^c|A \wedge C) = \Pi(B^c|A^c \wedge C) > 0$, *if T is strict.*

Theorem 4 and Theorem 5 allow to appreciate the difference of the same notion of independence under different notions of conditioning.

Under both rules of conditioning the characterization theorem of independence implies that our definition of independence is stronger than usual ones, in fact if $A \perp\!\!\!\perp B \mid C\,[\Pi]$ under a T-conditional possibility (or a T_{DP}-conditional possibility), then

$$\begin{aligned} \Pi(A|C) &= \max\{\Pi(A \wedge B|C), \Pi(A \wedge B^c|C)\} \\ &= \max\{T(\Pi(A|B \wedge C), \Pi(B|C)), T(\Pi(A|B^c \wedge C), \Pi(B^c|C))\} \\ &= \Pi(A|B \wedge C) \end{aligned} \quad (4)$$

and moreover

$$\Pi(A \wedge B|C) = T(\Pi(A|B \wedge C), \Pi(B|C)) = T(\Pi(A|C), \Pi(B|C)). \quad (5)$$

Equations (4) and (5) show that, for any continuous t-norm T, Definition 5 implies the classical notions of independence (see, e.g. de Cooman (1997b); Bouchon-Meunier et al. (2002)) and a rule that is a generalization of product rule that leads to interrelation for random variables.

The proposed notion of independence is not symmetric. Nevertheless, since there are just few separation criteria able to represent asymmetric independence models, symmetry is often required. From Theorem 4 we can obtain the corresponding result related to the symmetric property.

Corollary 1. *Let A and B be two logically independent events with respect to an event C. Consider a coherent T-conditional [T_{DP}-conditional] possibility (with T any continuous t-norm) Π on a set $\mathcal{G}$ containing $\mathcal{D} = \{A^*|B^* \wedge C, B^*|A^* \wedge C\}$, then $A \perp\!\!\!\perp B \,|\, C\,[\Pi]$ and $B \perp\!\!\!\perp A \,|\, C\,[\Pi]$ if and only if*

$$\Pi(A|B \wedge C) = \Pi(A|B^c \wedge C) \quad \textit{and} \quad \Pi(A^c|B \wedge C) = \Pi(A^c|B^c \wedge C)$$

and

$$\Pi(B|A \wedge C) = \Pi(B|A^c \wedge C) \quad \textit{and} \quad \Pi(B^c|A \wedge C) = \Pi(B^c|A^c \wedge C).$$

Corollary 2. *Let A and B be two logically independent events with respect to an event C. Consider a coherent T_{DP}-conditional possibility Π on a set $\mathcal{G}$ containing $\mathcal{D} = \{A^*|B^* \wedge C, B^*|A^* \wedge C\}$. If $A \perp\!\!\!\perp B \,|\, C\,[\Pi]$ then $B \perp\!\!\!\perp A \,|\, C\,[\Pi]$*

Proof. We prove the result for the minimum t-norm, since for other t-norms the proof goes along the same line.

If $A \perp\!\!\!\perp B \,|\, C\,[\Pi]$ one has either $\Pi(A^*|B^* \wedge C) = 1$ or $\Pi(A|B^* \wedge C) < 1$ (equivalently $\Pi(A^c|B^* \wedge C) < 1$) and $\Pi(B^*|A^* \wedge C) = 1$ from Theorem 5. Then the thesis follows. □

The symmetry property could be useful in order to handle situations of mutually independence of events and contrary to what happens under min-conditional possibility where symmetry can fail, while under T_{DP}-conditional possibility symmetry cannot fail.

4 Conditional independence for possibilistic variables

In this section we deal with a random vector $X_I = (X_i)_{i \in I}$ indexed by a finite set $I = \{1, \ldots, n\}$, where for each $i \in I$, the random variable X_i ranges in the finite set $\mathcal{X}_i$. Since some logical constraint (or structural zero) could be present among the variables we assume that X_I ranges in $\mathcal{X}_I \subseteq \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$.

As a consequence, for every $\emptyset \neq A \subseteq I$ with $A = \{i_1, \ldots, i_h\}$ we have a random vector $X_A = (X_i)_{i \in A}$ taking values in $\mathcal{X}_A \subseteq \mathcal{X}_{i_1} \times \cdots \times \mathcal{X}_{i_h}$. To simplify notation, we write (x_i) for the event $(X_i = x_i)$ and (x_i^c) for the event $(X_i \neq x_i)$ with $x_i \in \mathcal{X}_i$, while for every $\emptyset \neq A \subseteq I$, (x_A) stands for $(X_A = x_A)$ and (x_A^c) stands for $(X_A \neq x_A)$ with $x_A \in \mathcal{X}_A$.

The question now is the relation between different notions of independence for variables.

Definition 6. *Let A, B, C be mutually disjoint subsets of I with $A \neq \emptyset \neq B$ and Π a coherent T-conditional possibility [T_{DP}-conditional possibility] on $\mathcal{G}$ containing the set*

$$\begin{aligned} \mathcal{D} \;=\; & \{(x_A^*)|(x_B^*, x_C), (x_B^*)|(x_A^*, x_C) : \; \forall x_A \in \mathcal{X}_A, \forall x_B \in \mathcal{X}_B, \forall x_C \in \mathcal{X}_C, \\ & \qquad (x_B^*, x_C) \neq \emptyset \neq (x_A^*, x_C)\}. \end{aligned}$$

Then X_A **is conditionally independent of X_B given X_C under** *Π, in symbol $X_A \perp\!\!\!\perp X_B \,|\, X_C\,[\Pi]$, if for each $(x_A)|(x_B, x_C) \in \mathcal{D}$ it holds that $(x_A) \perp\!\!\!\perp (x_B) \,|\, (x_C)\,[\Pi]$.*

In the above definition, if we take $C = \emptyset$ and identify $(X_\emptyset = x_\emptyset) = \Omega$, we obtain a definition of unconditional independence between random vectors X_A and X_B, that we simply write as $X_A \perp\!\!\!\perp X_B\,[\Pi]$.

From equation (4) it follows that if $X_A \perp\!\!\!\perp X_B \mid X_C\,[\Pi]$ under a T-conditional possibility [T_{DP}-conditional possibility] then

$$\Pi(x_A|x_B, x_C) = \Pi(x_A|x_C)$$

for any $x_B \in \mathcal{X}_B$ and from equation (5)

$$\Pi(x_A, x_B|x_C) = T(\Pi(x_A|x_C), \Pi(x_B|x_C)).$$

This definition leads to an independence notion that is not necessarily symmetric, as already shown in Coletti and Vantaggi (2006) for $T = \min$ and in Ferracuti and Vantaggi (2006) for T a strict t-norms.

The next result allows us to appreciate the difference between independence under T-conditional possibility and T_{DP}-conditional possibility

Theorem 6. *Let T be the minimum. If a coherent T_{DP}-conditional possibility is such that $\Pi(x_A|x_B, x_C) = \Pi(x_A|x_C)$ varying $x_B \in \mathcal{X}_B$, for any $x_A \in \mathcal{X}_A$ and any $x_C \in \mathcal{X}_C$, then $X_A \perp\!\!\!\perp X_B \mid X_C\,[\Pi]$ if and only if one (and only one) of the following conditions holds:*

(a) $\Pi(x_A|x_C) = 1$ *for any* $x_A \in \mathcal{X}_A$ *and any* $x_C \in \mathcal{X}_C$*;*

(b) if $\Pi(x_A|x_C) < 1$ *then* $\Pi(x_B|x_A, x_C) = 1$ *for any* $x_B \in \mathcal{X}_B$*;*

Proof. If $\Pi(x_A|x_B, x_C) = a$ for any $x_B \in \mathcal{X}_B$, then $\Pi(x_A|x_C) = a$ and

$$T(\Pi(x_A|x_B^c, x_C), \Pi(x_B^c|x_C)) = \Pi(x_A, x_B^c|x_C) = \max_{x'_B \neq x_B} \{\Pi(x_A, x'_B|x_C)\} =$$

$$\max_{x'_B \neq x_B} \{T(\Pi(x_A|x'_B, x_C), \Pi(x'_B|x_C))\} = T\left(a, \max_{x'_B \neq x_B} \Pi(x'_B|x_C)\right) \leq a$$

simply by applying properties of t-norms.

Furthermore if $a = 1$ then by definition of T_{DP}-conditional possibility $\Pi(x_A|x_B^c, x_C) = 1$. Otherwise, for $a < 1$ again by definition of T_{DP}-conditional possibility one must have $\Pi(x_B^c|x_C) \geq a$ and then $\Pi(x_A|x_B^c, x_C) = a$.

Therefore, thesis follows from Theorem 5. □

A consequence of the above result and Theorem 5 is the following corollary.

Corollary 3. *Let T be the minimum. Under a coherent T_{DP}-conditional possibility Π if $X_A \perp\!\!\!\perp X_B \mid X_C\,[\Pi]$ then $X_B \perp\!\!\!\perp X_A \mid X_C\,[\Pi]$.*

The above results could be compared with Theorem 9 in Coletti and Vantaggi (2006) valid for T-conditional possibilities when the t-norm is the minimum.

Theorem 7. *Let T be a strict t-norm. If a coherent T_{DP}-conditional possibility is such that $\Pi(x_A|x_B, x_C) = \Pi(x_A|x_C)$ then $X_A \perp\!\!\!\perp X_B \mid X_C\,[\Pi]$ if and only if one (and only one) of the following conditions holds:*

(a) $\Pi(x_A|x_B) = 1$ *for any* $x_A \in \mathcal{X}_A$ *and any* $x_C \in \mathcal{X}_C$*;*

(b) if $\Pi(x_A|x_C) = 1$ *then* $\Pi(x_B|x_A, x_C) = 1$ *for any* $x_B \in \mathcal{X}_B$*;*

(c) if $\Pi(x_A|x_C) = \alpha \in (0, 1)$ *then* $\Pi(x_B|x_A, x_C) > 0$ *for any* $x_B \in \mathcal{X}_B$*.*

The proof goes along the same line of the previous one.

For strict t-norms the above results allow to compare this independence notion based on T_{DP}-conditional possibility with an analogous definition of independence based on zero-layers, instead of significant layers, based on T-conditional possibilities (Ferracuti and Vantaggi, 2006).

5 Conclusion

Conditioning and independence are the main concepts for updating information and for reasoning under hypotheses. The concept of conditioning cannot be relegated only to the role of restriction of the domain of possible events, when an event is occurred, but it is important to regard the conditioned and conditioning events as entities of the same kind, having in a certain moment a different role.

In this work the aim is to compare a concept of conditional independence based on two settings: the one of T-conditional possibility (Bouchon-Meunier et al., 2002; Coletti and Vantaggi, 2006; Ferracuti and Vantaggi, 2006) and the one that we call T_{DP}-conditional possibility, paying attention in particular to the cases the t-norm T is the minimum or strict. Since to handle significant layers can be not immediately understandable, we provided a characterization of conditional independence using only the values of the T-conditional possibility on $\mathcal{D}$. Due to the generality of continuous t-norms, this characterization needs to take into account many different situations.

In the paper we consider also T-conditional possibilities obtained through the minimum specificity principle, introduced by Dubois and Prade, regarded as a specific class of T-conditional possibilities. For them it is possible to introduce exactly the same notion of independence, that however has a different characterization in terms of the values of the T_{DP}-conditional possibility on $\mathcal{D}$.

References

S. Benferhat, K. Tabia, and K. Sedki. Jeffrey's rule of conditioning in a possibilistic framework. *Annals of Mathematics and Artificial Intelligence*, 66(3):185–202, 2011.

B. Bouchon-Meunier, G. Coletti, and C. Marsala. Independence and possibilistic conditioning. *Annals of Mathematics and Artificial Intelligence*, 35(1–4):107–123, 2002.

G. Coletti and D. Petturiti. Finitely maxitive T-conditional possibility theory: Coherence and extension. *International Journal of Approximate Reasoning*, 71:64–88, 2016.

G. Coletti and B. Vantaggi. Possibility theory: Conditional independence. *Fuzzy Sets and Systems*, 157(11):1491–1513, 2006.

G. Coletti and B. Vantaggi. T-conditional possibilities: Coherence and inference. *Fuzzy Sets and Systems*, 160(3):306–324, 2009.

G. Coletti, D. Petturiti, and B. Vantaggi. Independence in Possibility Theory under Different Triangular Norms. In L. van der Gaag, editor, *Symbolic and Quantitative Approaches to Reasoning ith Uncertainty*, pages 133–144, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

G. de Cooman. Possibility theory II: Conditional possibility. *International Journal of General Systems*, 25:325–351, 1997a.

G. de Cooman. Possibility theory III: Possibilistic independe. *International Journal of General Systems*, 25:353–371, 1997b.

B. de Finetti. Sull'impostazione assiomatica del calcolo delle probabilità. *Annali Università di Trieste*, 19:3–55, 1949. (English translation in: Chapter 5 of *Probability, Induction and Statistics*, Wiley, London, 1972).

D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty.* Plenum Press, New York and London, 1988.

L. Ferracuti and B. Vantaggi. Independence and conditional possibility for strictly monotone triangular norms: Research Articles. *Internationl Journal of Intelligent Systems*, 21(3):299–323, 2006.

N. Shilkret. Maxitive measure and integration. *Indagationes Mathematicae (Proceedings)*, 74:109–116, 1971.

The Generalized Running Intersection Property

Jane Ivy Coons[1,2], Carlotta Langer[3], and Michael Ruddy[4]

[1]St John's College, University of Oxford, United Kingdom
[2]Mathematical Institute, University of Oxford, United Kingdom
jane.coons@maths.ox.ac.uk
[3]Hamburg University of Technology, Hamburg, Germany
carlotta.langer@tuhh.de
[4]University of San Francisco, San Francisco, United States
mruddy@usfca.edu
[2]Mathematical Institute, University of Oxford, United Kingdom

Abstract

In this work we investigate *partition models*, the subset of log-linear models for which one can perform the iterative proportional scaling (IPS) algorithm to numerically compute the maximum likelihood estimate (MLE). Partition models include families of models such as hierarchical models and balanced, stratified staged trees. We define sufficient conditions, the Generalized Running Intersection Property (GRIP), on the matrix representation of a partition model for the IPS algorithm to always exactly produce the MLE in one cycle.

1 Introduction

The iterative proportional scaling (IPS) algorithm is a simple and efficient numerical algorithm for computing the maximum likelihood estimate (MLE). It can compute the MLE for certain families of log-linear statistical models, which we call *partition models*. Partition models include for instance hierarchical models, which have been heavily studied in connection with the IPS algorithm (Haberman, 1974; Lauritzen, 1996; Endo and Takemura, 2009; Xu et al., 2016).

From an information geometric perspective, calculating the MLE for a partition model can be described as projecting to linear families defined by the partitions of the matrix, that represents the model. Given a data vector and an estimate on the model, the IPS algorithm updates this estimate in each step by projecting onto a different linear family, converging towards the MLE. We say that the IPS algorithm has completed *one cycle* after it has iterated through each linear family exactly once.

In this work we are interested in the question, "When does the IPS algorithm exactly produce the MLE after one cycle?" We note that the matrix representation of a partition model heavily influences the performance of the IPS algorithm (see Example 2.5). In Haberman (1974) the author defines the Running Intersection Property (RIP) for hierarchical models, which gives sufficient conditions on the matrix representation of a hierarchical model so that the IPS algorithm always produces the MLE exactly in one cycle. An additional sufficient condition for one cycle convergence was proven in Vomel (1999). Drawing inspiration from the RIP, we define the Generalized Running Intersection Property (GRIP) on the matrix representations of partition models and show that it gives sufficient conditions for the IPS algorithm to always produce the MLE exactly in one cycle in Coons et al. (2022). We view rational partition models as the intersection of a toric variety with the probability simplex and in so doing, are able to employ methods from algebraic geometry, such as the toric fiber product, to understand the maximum likelihood estimates of these models. We additionally show that for hierarchical models the RIP is a special case of the GRIP. Finally, we are able to connect the GRIP to balanced, stratified staged trees.

In particular at each step of the IPS algorithm, the estimate is a rational function of the data vector, implying that in the case of one cycle convergence the MLE is a rational function. Models for which the MLE can always be described by a rational function of the data vector are called *rational* and are a subject of recent interest (Coons and Sullivant, 2021; Duarte et al., 2021; Huh and Sturmfels, 2014).

Our work gives sufficient conditions for a partition model to be rational. If the partition model has a parametrization which satisfies the GRIP, then the IPS converges in one cycle, the model is rational and one can read the MLE directly from the matrix of this parametrization. This work is structured as follows. In Section 2 we shortly introduce the mathematical background consisting of partition models, the IPS algoirthm and the running intersection property. This leads to the definition of the GRIP in Section 3 and we end with an outlook and discussion in Section 4.

2 Partition Models

In this section we define the the family of log-linear models that we call partition models. The IPS algorithm, defined in the next section, can be applied to these models and we will define the GRIP for them.

Consider an $n \times m$ matrix $A = (a_{ij})$ where $a_{ij} \in \mathbb{Z}_+$ and each column sum $\sum_{i=1}^{d} a_{ij}$ is equal. We assume throughout that A has the vector of all ones in its row span. Then the matrix A defines a homogeneous polynomial map $\phi_A : \mathbb{R}^n \to \mathbb{R}^m$ where

$$\phi_A(t_1, \ldots, t_n) = \left(\prod_{i=1}^{n} t_i^{a_{i1}}, \prod_{i=1}^{n} t_i^{a_{i2}}, \ldots, \prod_{i=1}^{n} t_i^{a_{im}} \right). \quad (1)$$

Let $\Delta_{m-1} \subset \mathbb{R}^m$ denote the open $(m-1)$-dimensional probability simplex.

Definition 2.1. The *log-linear* model associated to the integer matrix A, denoted $\mathcal{M}_A$ is the intersection of the Zariski closure of the image of ϕ_A with the open probability simplex; that is,

$$\mathcal{M}_A = \overline{\mathrm{Im}(\phi_A)} \cap \Delta_{m-1}.$$

The log-linear model $\mathcal{M}_A$ is the discrete exponential family whose sufficient statistics are given by the rows of A, hence it can also be written as

$$\mathcal{M}_A = \{p \in \Delta_{m-1} | \log p \in rowspan(A)\}.$$

The models we are interested in have a matrix A that can be structured in the following way. Suppose each column sum of A is equal to k, meaning $\sum_{i=1}^{n} a_{ij} = k$ for all $1 \leq j \leq n$. Then one can group the rows of A into matrices A^ℓ with column equal to one where $1 \leq \ell \leq k$. In each matrix A^ℓ, exactly one row of A^ℓ has a non-zero entry for each column of A and hence encodes a partition of the state space $\mathcal{X}$. For this reason we refer to matrices A^ℓ, $1 \leq \ell \leq k$ as the *partitions* of A.

Definition 2.2. A matrix A gives rise to a **partition model** $\mathcal{M}_A$ if the map in (1) is a homogeneous, multi-linear monomial map.

Let A^1A^2 denote the matrix obtained by stacking A^1 above A^2; that is,

$$A^1A^2 := \begin{bmatrix} A^1 \\ \hline A^2 \end{bmatrix}. \tag{2}$$

From the partitions of A we can build a new matrix, by stacking the partitions as defined above $A^{1,\dots,k} = A^1A^2 \dots A^k$ where α_i^ℓ denotes the i-th row of the ℓ-th partition, as illustrated in Example 2.4. Although there are many ways to arrange the rows of A to build $A^{1,\dots,k}$, since they all have the collection of rows as A they clearly have the same row-span, and hence define the same toric model $\mathcal{M}_A$ as A. However, as we show in the next subsection, a different representation of the same model may affect the convergence of the iterative proportional scaling algorithm. Without loss of generality we assume A is of the above form.

Definition 2.3. The **index set** of α_i^ℓ, denoted I_i^ℓ, is the set of indices $j \in \{1, \dots, m\}$ such that the j-th entry of α_i^ℓ is one.

For a fixed partition A^ℓ and index $j \in \{1, \dots, m\}$ there is exactly one row α_i^ℓ such that j lies in its index set. We can define a function $\mathcal{S}(\ell, j) \in \{1, \dots, n_l\}$ where $\mathcal{S}(\ell, j)$ is the index such that $j \in I_{\mathcal{S}(\ell,j)}^\ell$. Then $\alpha_{\mathcal{S}(\ell,j)}^\ell$ is the row of A^ℓ where j lies in its index set.

Example 2.4. Let us now consider the matrix A on the right. In this example $m = 7, n_1 = 2$ and $n_2 = 2$. Then the index set $I_2^2 = \{4, 5, 6, 7\}$ and $\mathcal{S}(1, 3) = 2$.

$$j \in \{\, 1,\ 2,\ \cdots \qquad \cdots\ 6,\ 7 \,\}$$

$$A = \left(\begin{array}{ccccccc} 1 & \cdot & \cdot & 1 & 1 & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot & \cdot & 1 & 1 \\ \hline 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 \end{array}\right) \begin{array}{c} \alpha_1^1 \\ \alpha_2^1 \\ \hline \alpha_1^2 \\ \alpha_2^2 \end{array} \begin{array}{l} \left.\begin{array}{c} \\ \\ \end{array}\right\} A^1 \\ \left.\begin{array}{c} \\ \\ \end{array}\right\} A^2 \end{array}$$

We are able to apply the iterative proportional scaling algorithm as described in the next section to a partition model $\mathcal{M}_A$ in order to estimate the maximum likelihood of a data vector in $\mathbb{R}^m_+$.

2.1 Iterative Proportional Scaling

The iterative proportional scaling (IPS) algorithm is a method to calculate the maximum likelihood estimation of a normalized data vector d with respect to the model $\mathcal{M}_A$. This algorithm first appeared in the statistics literature in (Deming and Stephan, 1940) and was further analyzed for example in (Csiszár, 1975). Before we discuss the algorithm in more detail, we first define the maximum likelihood method.

Maximum likelihood estimation is a way to find an element in a statistical model, $\mathcal{M}_A$ that fits the observed data best. Let d be the empirical distribution of the data. Then the maximum likelihood estimator is given by

$$p^* = \arg\max_{p \in \mathcal{M}_A} \sum_{j=1}^{m} d_j \, log \, p_j .$$

Here we use the log-likelihood function. More details can be found in for example Sullivant (2018) and Drton et al. (2008). If $p^\star$ is rational for every d, then we say that $\mathcal{M}_A$ has rational MLE.

Now we will define the steps of the IPS. The starting point is the uniform distribution $p^0 = (\frac{1}{n}, \dots, \frac{1}{n})$, the ℓth-step of the algorithm is then defined as

$$p^\ell = p^{\ell-1} * \frac{A^i d}{A^i p^{\ell-1}}$$

for $i = \ell \bmod k$.

Every step is an information projection to the linear family

$$L^i = \{p \in \Delta_{m-1} | \, A^i p = A^i d\}.$$

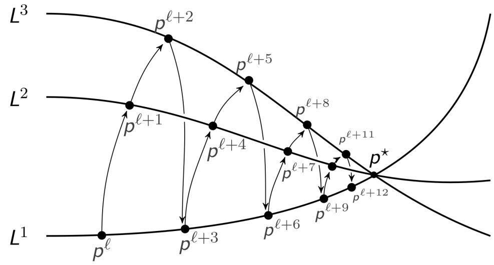


Figure 1: Sketch of the IPS.

This iterative process is sketched in Figure 1. Note that each step of the IPS produces a rational function. A proof of the convergence can be found in (Csiszár and Shields, 2004) Theorem 5.1. The next example demonstrates that the rate of convergence heavily depends on the chosen representation.

Example 2.5. Here we consider the two matrices A and $\tilde{A}$ depicted below.

$$A = \begin{pmatrix} 1 & 1 & \cdot \\ \cdot & \cdot & 1 \\ \hdashline 1 & \cdot & \cdot \\ \cdot & 1 & 1 \end{pmatrix} \qquad \tilde{A} = \begin{pmatrix} 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \end{pmatrix}$$

Both matrices have full rowspan and $\mathcal{M}_A = \mathcal{M}_{\tilde{A}}$ holds. Although they represent the same model, the convergence of the IPS algorithm is heavily influenced by the chosen representation. Using matrix $\tilde{A}$ the IPS algorithm converges in one step to the MLE $p^\star = d$.

The first two steps of the IPS algorithm on A with normalized data vector $d = (d_1, d_2, d_3)$ results in

$$p^0 = \left(\frac{1}{2}(d_1 + d_2),\ \frac{1}{2}(d_1 + d_2), d_3\right)$$
$$p^1 = \left(d_1,\ \frac{1}{2}(d_1 + d_2)\frac{(d_2 + d_3)}{\frac{1}{2}(d_1 + d_2) + d_3}, d_3\frac{(d_2 + d_3)}{\frac{1}{2}(d_1 + d_2) + d_3}\right)$$

In general the different projections have the following form:

$$p^k = \left(d_1,\qquad p_2^{k-1}\frac{(d_2 + d_3)}{p_2^{k-1} + d_3}, d_3\frac{(d_2 + d_3)}{p_2^{k-1} + d_3}\right), \quad k \text{ odd} \tag{3}$$
$$p^k = \left(d_1\frac{(d_1 + d_2)}{d_1 + p_2^{k-1}}, p_2^{k-1}\frac{(d_1 + d_2)}{d_1 + p_2^{k-1}}, d_3\right), \quad k \text{ even} \tag{4}$$

Assuming that there exists an index k such that the second entry of p^k is exactly d_2 in (3) and (4) at the same time, leads to $d_2 = p_2^{k-1}$. Hence the IPS can only result in the exact MLE, if $d_1 = d_2$.

In a practical evaluation with 20 000 random input distributions the arithmetic mean of the iteration steps taken to get a step size smaller than 10^{-8} was $113,4767$ with a minimum value of 8 and a maximum value of 287 478. Note that in case of $\tilde{A}$ the necessary iteration steps are only 1.

This example demonstrates the importance of choosing a convenient representation of a partition model. The next section introduces the running intersection property, that defines conditions that lead to a representation guaranteeing one-cycle convergence.

2.2 The Running Intersection Property

A hierarchical model is a partition model for which the interaction structure between the columns can be described by a simplicial complex. We will briefly introduce hierarchical models, more details can be found in (Sullivant, 2018; Coons et al., 2022).

Definition 2.6. Let 2^J be the powerset of $J = \{1, \ldots, l\}$, meaning the set of all the subsets of J. A **simplicial complex** Γ with the ground set J is a subset of 2^J with the property that if $F \in \Gamma$ and $F' \subset F$, then $F' \in \Gamma$. The elements of a simplicial complex are the faces of Γ and all inclusion maximal faces are called facets.

Here we define the matrix A_Γ corresponding to a simplicial complex Γ with facets $\{F_1, \ldots F_r\}$. The columns of the matrix A_Γ are indexed by subsets of $[n] = \{1, \ldots, n\}$.

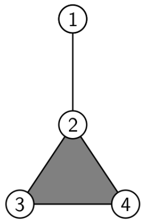
Note that this is the same as being indexed by 0/1 strings of length n by taking each $S \subset [n]$ to be the the set of positions in the string equal to 1.

The rows of A_Γ are divided into blocks $A^1, \ldots, A^k$ each of size $2^{|F_i|}$, $i \in \{1, \ldots, r\}$. The rows of block A^i are of the form a^i_S where $S \subset F_i$ with

$$a^i_S(T) = \begin{cases} 1, & \text{if } T \cap F_i = S \\ 0, & \text{otherwise,} \end{cases}$$

for each $T \subset [n]$. Such a matrix is given in Example 2.7.

Example 2.7. Let Γ consist of two facets $\{1, 2\}$ and $\{2, 3, 4\}$, then one can draw the graphical representation of the associated hierarchical model as depicted on the right. The matrix A_Γ consists of two partitions, corresponding to the two facets.



$$A_\Gamma = \left(\begin{array}{cccccccccccccccc}
1 & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 \\
\hline
1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1
\end{array}\right)$$

Figure 2: Hierarchical model and its corresponding matrix.

Now we will introduce a special ordering to the facets of a simplicial complex.

Definition 2.8. Let $E = \{F_1, \ldots, F_s\}, s \in \mathbb{N}$ be an ordering of the facets of a simplicial complex Γ. Then this ordering satisfies the running intersection property (RIP), if for each $r \in \{1, \ldots, s\}$ there exists a k_r such that

$$\left(\bigcup_{k=1}^{r} F_k \right) \cap F_{r+1} = F_{k_r} \cap F_{r+1}$$

Theorem 2.9. *Suppose there exists an ordering such that the simplicial complex Γ satisfies the RIP and the initial distribution lies in the model $\mathcal{M}_\Gamma$, then the IPS algorithm converges in one cycle.*

Proof. This theorem is proven in Theorem 5.3 in Haberman (1974). □

3 The Generalized Running Intersection Property

In this section, we define the generalized running intersection property, or GRIP. We derive a formula for the maximum likelihood estimate of a model that satisifies the GRIP and use this to show that iterative proportional scaling exhibits one-cycle convergence on models that satisfy the GRIP.

In order to define the GRIP, we first define the floret condition. The word choice of "floret" in the definition below is deliberately suggestive of the terminology for staged trees. In Coons et al. (2022) we highlight the connection between the GRIP and staged trees in more detail and show that we require that the tree associated to the partition matrix is staged. In Example 3.8 we depict a matrix A together with its associated tree.

Definition 3.1. Let B and C be two partition matrices with the same number of columns and with rows α_u^B and α_v^C. The matrices B and C satisfy the **floret condition** if for every two rows of B, α_u^B and $\alpha_{u'}^B$, the sets of rows of C that are connected to α_u^B and $\alpha_{u'}^B$ are disjoint or equal. In this case, the set of rows of B connected to a row α_v^C is called a **floret** of B and the set of rows of C connected to a row α_u^B is a **floret** of C.

Example 3.2. The matrix on the right does not satisfy the floret condition, since the first row α_1^1 is connected to both rows in the second partition and the second row α_1^2 is only connected to α_1^2.

$$\begin{pmatrix} 1 & \cdot & \cdot & 1 & 1 & 1 & 1 \\ \cdot & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \hdashline 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 \end{pmatrix}$$

For a set of partition matrices $A^1, \dots, A^\ell$ with m columns, let $B = \biguplus_{n=1}^{\ell} A^n$. Let $C = A^{\ell+1}$ have γ rows and m columns. Then the matrix BC has m columns of the form $[e_u \ e_v]^T$ where $u \in [\beta]$ and $v \in [\gamma]$. Let ω_{uv} denote the number of columns of C of the form $[e_u \ e_v]^T$. Then the columns of BC can be indexed by triples of the form (u, v, s) for $s \in \{0, \dots, \omega_{uv} - 1\}$ where column (u, v, s) is the $(s+1)$st column of BC of the form $[e_u \ e_v]^T$.

Definition 3.3. Suppose that B and C satisfy the floret condition with f florets, $\mathcal{F}_1^C, \dots, \mathcal{F}_f^C$. We define the matrix $B \Cap C$ to be the $f \times m$ matrix whose columns are indexed by the triples (u, v, s) such that the (u, v, s) entry of the tth row of $B \Cap C$ is equal to 1 if $t = t(u, v)$ and 0 otherwise. Note that any two columns with indices (u, v, s) and (u, v, s') are identical.

In other words, the columns of $B \Cap C$ are indicator vectors for the florets that each column's non-zero rows belong to. Note that $B \Cap C$ is only defined when B and C satisfy the floret condition.

Example 3.4. Performing the above defined operation on the first two partitions of A from Example 3.8 leads to $A^1 \Cap A^2 = (1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1)$.

Definition 3.5. Let B and C be two partition matrices with the same number of columns and with rows α_i^B and $\alpha_{i'}^C$. Two rows α_i^B and $\alpha_{i'}^C$ are **connected** if their supports intersect nontrivially; that is, if $I_i^B \cap I_{i'}^C \neq \emptyset$.

For a matrix to satisfy the GRIP, we require that rows from one partition to the next are connected in way that satisfies certain conditions.

Definition 3.6. For a matrix A of the form (2), the **column weight** of the j-th column is number of times the column is repeated and is denoted $c_j \in \mathbb{Z}_+$.

The following definition is motivated by the way column weights behave under steps of the IPS algorithm.

Definition 3.7. For a matrix A of the form (2), define c_j^ℓ as the j-th column weight for the matrix obtained by only considering the first ℓ partitions of A. Then A is **well-connected** if for any row vector α_i^ℓ, where $\ell > 1$, we have that $(c_j^\ell / c_j^{\ell-1}) = (c_{j'}^\ell / c_{j'}^{\ell-1})$ for all $j, j' \in I_i^\ell$. We call this quantity the **connection ratio** for α_i^ℓ denote this quantity as C_i^ℓ with the convention that $C_i^1 = |I_i^1|$.

Example 3.8. Consider the matrix A on the right. The connection ratios for the rows of A are depicted on the right of the matrix. We note that A it is well-connected. Below the matrix we have the associated staged tree representation of the matrix. The vertical dotted lines show which columns of the matrix A relate to the third-level florets of the staged tree model.

$$\mathcal{F}_{v_0}, \quad \mathcal{F}_{v_{s_0}} = \mathcal{F}_{v_{s_1}},$$
$$\mathcal{F}_{v_{s_0 t_0}} = \mathcal{F}_{v_{s_1 t_0}}, \quad \mathcal{F}_{v_{s_0 t_1}} = \mathcal{F}_{v_{s_1 t_1}}.$$

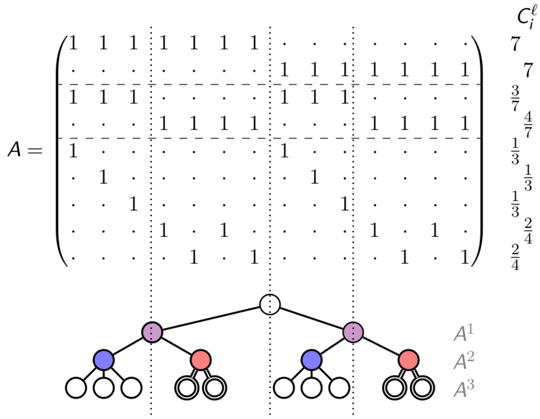


Figure 3: Matrix A with its connection ratios and associated tree.

See (Ananiadi and Duarte, 2021, Section 2) for an introduction to staged trees and the corresponding definitions.

Let $A^1, \dots, A^\ell$ be a set of partition matrices with m columns and $A^{1,\dots,\ell} = A^1 A^2 \dots A^\ell$. Let β be number of *distinct* columns of $A^{1,\dots,\ell}$. Define a labeling of the columns of $A^{1,\dots,\ell}$, $\lambda : [m] \to [\beta]$ such that $\lambda(j) = \lambda(j')$ if and only if the jth and j'th columns of $A^{1,\dots,\ell}$ are equal.

Definition 3.9. We define the partition matrix $B = \Cup_{n=1}^{\ell} A^n$ to be the $\beta \times m$ matrix with jth column $e_{\lambda(j)}$. Since the labeling λ of the columns of $A^{1,\dots,\ell}$ simply permutes the rows of $\Cup_{n=1}^{\ell} A^n$, we omit the specification of λ from this notation.

Example 3.10. The matrix $B = A^1 \Cup A^2$ consisting of the partial matrices of A from Example 3.8 results in

$$A^1 \Cup A^2 = \begin{pmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 \end{pmatrix}$$

Definition 3.11. Let $A^1, \dots, A^k$ be partition matrices. For each ℓ, let B_ℓ denote $\Cup_{n=1}^{\ell} A^n$. Then $A^{1,\dots,k}$ satisfies the **generalized running intersection property**, or **GRIP** if for each $1 \le \ell \le k-1$,

1. the matrix $B_\ell A^{\ell+1}$ is well-connected,
2. $B_\ell A^{\ell+1}$ satisfies the floret condition, and
3. the rows of $B_\ell \Cap A^{\ell+1}$ lie in the rowspan of $A^{1,\dots,\ell}$.

If a matrix satisfies the GRIP, then we are able to give a formula for the MLE.

Corollary 3.12. *Let A be a partition matrix with k partitions that satisfies the GRIP. Then the MLE p^* of d has as its jth coordinate function:*

$$p_j^* = \frac{1}{c_j}\left(\prod_{\ell=1}^{k} \frac{\alpha^{\ell}_{\mathcal{S}(\ell,j)}(d)}{\sum\limits_{\alpha_i^{\ell}\in\mathcal{F}^{\ell}_{\mathcal{S}(\ell,j)}} \alpha_i^{\ell}(d)}\right). \tag{5}$$

The proof of the above result is given in (Coons et al., 2022). Using this, we are able to prove the one-cycle convergence.

Theorem 3.13. *If A is a partition matrix with k partitions that satisfies the GRIP given in Definition 3.11 then the IPS algorithm results in the MLE after one cycle.*

In order to prove this result, we employ methods from algebraic geometry, such as the toric fiber product, to understand the maximum likelihood estimates of these models, given in Corollary 3.12.

Example 3.14. Here we apply the IPS to a matrix A, that satisfies the GRIP.

$$A = \left(\begin{array}{cccccccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\hdashline
1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & 1 & 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 \\
\hdashline
1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & 1 & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & 1 & \cdot \\
\cdot & \cdot & \cdot & \cdot & 1 & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & 1
\end{array}\right)
\begin{array}{c}
\alpha_1^1 \\ \alpha_2^1 \\ \hdashline \alpha_1^2 \\ \alpha_2^2 \\ \hdashline \alpha_1^3 \\ \alpha_2^3 \\ \alpha_3^3 \\ \alpha_4^3 \\ \alpha_5^3
\end{array}$$

Let $d = (d_1, \dots, d_{14})$ be the normalized data vector. Projecting to the first partition of A leads to:

$$p_1^0 = \cdots = p_7^0 = \frac{1}{7}\,\alpha_1^1(d), \qquad p_8^0 = \cdots = p_{14}^0 = \frac{1}{7}\,\alpha_2^1(d)$$

The second step of the algorithm results in four different types of indices. These are given by the different rows in the matrix $A^1 \uplus A^2$, indicated by the different dashed and dotted lines below the matrix.

$$A^1 \Cap A^2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \qquad \alpha_1^{1\Cap 2}$$

$$A^1 \uplus A^2 = \begin{pmatrix}
1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & 1 & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1
\end{pmatrix}
\begin{array}{c} \alpha_1^{1\uplus 2} \\ \alpha_2^{1\uplus 2} \\ \alpha_3^{1\uplus 2} \\ \alpha_4^{1\uplus 2} \end{array}$$

Note that since $A^1 \Cup A^2$ consists only of the ones vector and d is normalized, we have $\alpha_1^{1\Cap 2}(d) = 1$. Therefore the second projection simplifies in the following way:

$$
\begin{aligned}
p_1^1 = p_2^1 = p_3^1 = \quad & \frac{1}{C_1^1}\alpha_1^1(d)\frac{\alpha_1^2(d)}{C_1^2\,\alpha_1^{1\Cap 2}(d)} = \frac{1}{3}\alpha_1^1(d)\alpha_1^2(d)\\
p_4^1 = p_5^1 = p_6^1 = p_7^1 = \quad & \frac{1}{C_1^1}\alpha_1^1(d)\frac{\alpha_2^2(d)}{C_2^2\,\alpha_1^{1\Cap 2}(d)} = \frac{1}{4}\alpha_1^1(d)\alpha_2^2(d)\\
p_8^1 = p_9^1 = p_{10}^1 = \quad & \frac{1}{C_2^1}\alpha_2^1(d)\frac{\alpha_1^2(d)}{C_1^2\,\alpha_1^{1\Cap 2}(d)} = \frac{1}{3}\alpha_2^1(d)\alpha_1^2(d)\\
p_{11}^1 = p_{12}^1 = p_{13}^1 = p_{14}^1 = \quad & \frac{1}{C_2^1}\alpha_2^1(d)\frac{\alpha_2^2(d)}{C_2^2\,\alpha_1^{1\Cap 2}(d)} = \frac{1}{4}\alpha_2^1(d)\alpha_2^2(d)
\end{aligned}
$$

For the last projection we only demonstrate four indices as an example.

$$
(A^1 \Cup A^2) \Cap A^3 = \begin{pmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 & 1 & 1 & 1 \end{pmatrix} \qquad \begin{matrix} \alpha_1^{(1\Cup 2)\Cap 3} \\ \alpha_2^{(1\Cup 2)\Cap 3} \end{matrix}
$$

In this case $A^2 = (A^1 \Cup A^2) \Cap A^3$, hence $\alpha_i^2(d) = \alpha_i^{(1\Cup 2)\Cap 3}(d)$. In order to visualize the structure created by the IPS, we also display the general formula of p_j^2 without simplifications

$$
\begin{aligned}
p_1^2 &= \frac{1}{C_1^1}\alpha_1^1(d)\frac{\alpha_1^2(d)}{C_1^2\,\alpha_1^{1\Cap 2}(d)}\frac{\alpha_1^3(d)}{C_1^3\alpha_1^{(1\Cup 2)\Cap 3}(d)} = \frac{1}{7}\alpha_1^1(d)\frac{\alpha_1^2(d)}{\frac{3}{7}\,\alpha_1^{1\Cap 2}(d)}\frac{\alpha_1^3(d)}{\frac{1}{3}\alpha_1^{(1\Cup 2)\Cap 3}(d)} = \alpha_1^1(d)\alpha_1^3(d)\\
p_4^2 &= \frac{1}{C_1^1}\alpha_1^1(d)\frac{\alpha_2^2(d)}{C_2^2\,\alpha_1^{1\Cap 2}(d)}\frac{\alpha_4^3(d)}{C_4^3\alpha_2^{(1\Cup 2)\Cap 3}(d)} = \frac{1}{7}\alpha_1^1(d)\frac{\alpha_2^2(d)}{\frac{4}{7}\,\alpha_1^{1\Cap 2}(d)}\frac{\alpha_4^3}{\frac{2}{4}\alpha_2^{(1\Cup 2)\Cap 3}(d)} = \frac{1}{2}\alpha_1^1(d)\alpha_4^3(d)\\
p_8^2 &= \frac{1}{C_2^1}\alpha_2^1(d)\frac{\alpha_1^2(d)}{C_1^2\,\alpha_1^{1\Cap 2}(d)}\frac{\alpha_1^3(d)}{C_1^3\alpha_1^{(1\Cup 2)\Cap 3}(d)} = \frac{1}{7}\alpha_2^1(d)\frac{\alpha_1^2(d)}{\frac{3}{7}\,\alpha_1^{1\Cap 2}(d)}\frac{\alpha_1^3(d)}{\frac{1}{3}\alpha_1^{(1\Cup 2)\Cap 3}(d)} = \alpha_2^1(d)\alpha_1^3(d)\\
p_{11}^2 &= \frac{1}{C_2^1}\alpha_2^1(d)\frac{\alpha_2^2(d)}{C_2^2\,\alpha_1^{1\Cap 2}(d)}\frac{\alpha_4^3(d)}{C_4^3\alpha_2^{(1\Cup 2)\Cap 3}(d)} = \frac{1}{7}\alpha_2^1(d)\frac{\alpha_2^2(d)}{\frac{4}{7}\,\alpha_1^{1\Cap 2}(d)}\frac{\alpha_4^3}{\frac{2}{4}\alpha_2^{(1\Cup 2)\Cap 3}(d)} = \frac{1}{2}\alpha_1^1(d)\alpha_4^3(d)
\end{aligned}
$$

For the indices 4 and 11 there exists two identical columns, respectively. Hence the connection ratios to not add up to one, but to $\frac{1}{2}$.

Remark 3.15. For a partition matrix A with $k = 2$, it is possible to show that if the IPS algorithm produces the MLE in one cycle, then A satisfies the GRIP. This relies on being able to explicitly write the coordinate functions of the MLE as a product of linear forms that must satisfy the Horn uniformization (Duarte et al., 2021; Huh and Sturmfels, 2014). While the above counter-example implies that this is not the case for $k = 3$, it would be interesting to investigate whether requiring IPS to produce the MLE of $A^{1,\dots,\ell}$ at the ℓ-th step is sufficient to guarantee that A satisfies the GRIP.

4 Outlook and Discussion

This work gives a short introduction to the GRIP, in (Coons et al., 2022) we additionally draw connections between the GRIP and balanced stratified staged trees as well as toric fiber products. Figure 4 shows the main results that we prove in (Coons et al., 2022).

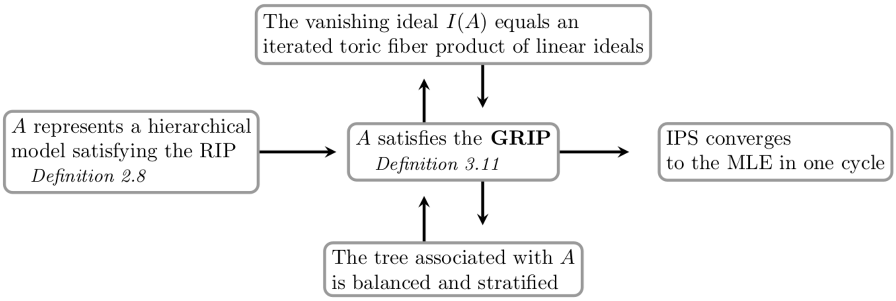


Figure 4: Sketch of the main results shown in (Coons et al., 2022)

We conclude with a discussion of some natural questions that arise from this work.

Question 1. For a partition matrix A with k partitions, does requiring that the IPS algorithm produces the MLE for $A^{1,\dots,\ell}$ at each step imply that A must satisfy the GRIP?

In Remark 3.15, we claim that IPS producing the MLE in 2 steps for a partition matrix with 2 partitions is enough to guarantee that the GRIP is satisfied. While the GRIP is sufficient for the algorithm to produce the MLE for $A^{1,\dots,\ell}$ at each step, it is possible that the reverse logical direction holds.

Question 2. Can one always find a representation of a log-linear partition model with rational MLE such that it satisfies the GRIP?

It is clear that the matrix representation of a particular model affects the IPS algorithm. In (Coons and Sullivant, 2021), the authors show sufficient conditions for 2-way quasi-independence models to have rational MLE. Since k-way quasi-independence models are just partition models without repeated columns, it would be interesting to see if these results can be used to show that every such matrix has a representation satisfying the GRIP. Indeed our preliminary results indicate that this is true, but we do not include a formal proof.

Finally there are many results and questions related to the convergence of the generalized IPS algorithm on log-linear models and its connections to tools from algebraic statistics (see (Améndola et al., 2021, Sec. 5) and (Drton et al., 2008, Sec. 7.3)). We feel that our work falls adjacent to this line of inquiry and that it would be interesting to investigate whether tools that have recently produced results in this area can connect to our work.

References

C. Améndola, K. Kohn, P. Reichenbach, and A. Seigal. Toric invariant theory for maximum likelihood estimation in log-linear models. *Algebraic Statistics*, 10 2021.

L. Ananiadi and E. Duarte. Gröbner bases for staged trees. *Algebraic Statistics*, 12(1): 1–20, Apr. 2021. doi: 10.2140/astat.2021.12.1.

J. I. Coons and S. Sullivant. Quasi-independence models with rational maximum likelihood estimator. *Journal of Symbolic Computation*, 104:917–941, May 2021. doi: 10.1016/j.jsc.2020.10.006.

J. I. Coons, C. Langer, and M. Ruddy. Rational partition models under iterative proportional scaling. Preprint, 2022. URL `https://mgruddy.github.io/partitionmodels`.

I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3, 1975.

I. Csiszár and P. C. Shields. *Information Theory and Statistics: A Tutorial*. Foundations and Trends in Communications and Information Theory. Now Publishers Inc, 2004.

W. E. Deming and F. F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 1940.

M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on algebraic statistics*, volume 39. Springer Science & Business Media, 2008.

E. Duarte, O. Marigliano, and B. Sturmfels. Discrete statistical models with rational maximum likelihood estimator. *Bernoulli*, 27(1), Feb. 2021. doi: 10.3150/20-bej1231.

Y. Endo and A. Takemura. Iterative proportional scaling via decomposable submodels for contingency tables. *Comput. Statist. Data Anal.*, 53(4):966–978, 2009. doi: 10.1016/j.csda.2008.11.013. URL `https://doi.org/10.1016/j.csda.2008.11.013`.

S. J. Haberman. *The Analysis of Frequency Data*. University of Chicago Press, 1974.

J. Huh and B. Sturmfels. Likelihood geometry. In *Lecture Notes in Mathematics*, pages 63–117. Springer International Publishing, 2014. doi: 10.1007/978-3-319-04870-3_3. URL `https://doi.org/10.1007/978-3-319-04870-3_3`.

S. L. Lauritzen. *Graphical Models*. Clarendon Press, 1996.

S. Sullivant. *Algebraic Statistics*, volume 194. American Mathematical Soc., 2018.

J. i. Vomel. Methods of probabilistic knowledge integration, 1999. URL `http://staff.utia.cas.cz/vomlel/prace.pdf`.

P.-F. Xu, J. Sun, and N. Shan. Local computations of the iterative proportional scaling procedure for hierarchical models. *Comput. Statist. Data Anal.*, 95:17–23, 2016. ISSN 0167-9473. doi: 10.1016/j.csda.2015.10.009. URL `https://doi.org/10.1016/j.csda.2015.10.009`.

Classes of Conflictness / Non-Conflictness of Belief Functions

Milan Daniel[1] and Václav Kratochvíl[2]

[1]Institute of Computer Sciences, Czech Academy of Sciences, Prague, Czech Republic
milan.daniel@cs.cas.cz
[2]Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech Republic
velorex@utia.cas.cz

Abstract

Theoretic, descriptive and experimental analysis and description of classes of conflictness, non-conflictness and of conflict hiddeness of belief functions. Theoretic extension of theory of hidden conflicts. Idea of catalogue of belief structures.

1 Introduction

As discussed in [2], the weight of conflict according to the classic Shafer's definition [13] using $m_{\bigcirc\!\!\!\!\cap}(\emptyset)$ is frequently higher than the expected value of conflict even for the partially conflicting belief functions (BFs). On the other hand, a positive value of a conflict (here we have in mind the conflict between BFs based on their non-conflicting parts [3, 4]) was observed even in a situation when $m_{\bigcirc\!\!\!\!\cap}(\emptyset)$ equals zero.

This observation led to the definition of several degrees (up to cardinality of frame of discernment) of hidden conflicts [6, 8], later compared with alternative shades of conflict [12] in [7]. In one-to-one relation to different degrees of conflict hiddenness, there are corresponding classes of non-conflictness [5]. And it is precisely the content of this paper — to explore and analyze different classes of belief functions concerning the hidden conflict.

This study covers a theoretical, descriptive, and a experimental approaches. The first one analyses the definitions and conditions of particular degrees of non-conflictness, resulting in a theoretic characterization of specific classes of non-conflicting BFs in various degrees. The other approaches characterize types of BFs, intending to describe and catalog different structures of BFs concerning hidden conflict or non-conflictness. Due to the complexity of BFs and sets/structures of their focal elements, this approach brings detailed results on small frames of discernment (two and three-element frames Ω_2, Ω_3) and rougher results for larger frames and a general n-element frame Ω_n.

The idea of going through all possible structures of belief functions on different frames of discernment and arranging them in a catalogue was inspired by a similar work [15] on a system of min-balanced systems known from game theory. Indeed, both the system of coalitions and the structure of a belief function is a set of sets. However, unlike game theory, the structure of belief functions is much richer — any restrictive rule does not limit it. For this reason, the resulting catalogue is quite extensive. Therefore it may be appropriate to limit it to some structurally limited subclass of belief functions, such as consonant belief functions in the future. An example of a catalogue for a particular class of min-balanced systems can be found at http://gogo.utia.cas.cz/indecomposable-min-semi-balanced-catalogue. We plan to create a similar catalogue, however, it is not finished at the time of writing this paper.

2 Basic Notions

This section will recall some basic notations needed in this paper.

Assume a finite frame of discernment Ω with elements denoted by lower-case letters from Latin alphabet $a, b, c, \ldots$ and their sets by capital letters. $A = \{a, b\}$. To simplify the notation, we abbreviate $\{a, b\}$ with ab. In the case of $|\Omega| = n$, we will highlight this fact using a subscript as Ω_n. $\mathcal{P}(\Omega) = \{X | X \subseteq \Omega\}$ is a *power-set* of Ω. $\mathcal{P}(\Omega)$ is often denoted also by 2^{Ω}, e.g., in [12].

A *basic belief assignment (bba)* is a mapping $m : \mathcal{P}(\Omega) \longrightarrow [0, 1]$ such that $\sum_{A \subseteq \Omega} m(A) = 1$. The values of the bba are called *basic belief masses (bbm)*. $m(\emptyset) = 0$ is usually assumed. We sometimes speak about m as of a mass function.

There are other equivalent representations of m: A *belief function (BF)* is a mapping $Bel : \mathcal{P}(\Omega) \longrightarrow [0, 1]$, $Bel(A) = \sum_{\emptyset \neq X \subseteq A} m(X)$. Because there is a unique correspondence between m and corresponding Bel we often speak about m as of a belief function.

Let m be a belief function defined on Ω and $A \subseteq \Omega$. If $m(A) > 0$ we say A is a *focal element* of m. The set of focal elements is denoted by $\mathcal{F}_m$ (or simply $\mathcal{F}$ for short) and we call it a *structure* of m. We say that a focal element $X \in \mathcal{F}$ is proper if $X \neq \Omega$. In the case of $m_{vac}(\Omega) = 1$ we speak about the *vacuous BF* (VBF) and about *a non-vacuous BF* otherwise. We speak about *consistent* BF if all focal elements have a non-empty intersection. If focal elements are nested, we speak about consonant BF.

The (non-normalized) conjunctive rule of combination $\odot$, see e.g. [14], is defined by:

$$(m_1 \odot m_2)(A) = \sum_{X \cap Y = A; X, Y \subseteq \Omega} m_1(X) m_2(Y)$$

for any $A \subseteq \Omega$. $\kappa = \sum_{X \cap Y = \emptyset; X, Y \subseteq \Omega} m_1(X) m_2(Y)$ is usually considered to represent a conflict of respective belief functions when $\kappa > 0$. By normalization of $m_{12} = m_1 \odot m_2$ we obtain Dempster's rule, see [13]. To simplify formulas, we often use $\bigodot_1^3 m = m \odot m \odot m$, and also $\bigodot_1^k (m_1 \odot m_2) = (m_1 \odot m_2) \odot \ldots \odot (m_1 \odot m_2)$, where $(m_1 \odot m_2)$ is repeated k-times.

3 Hidden Conflicts and Internal Hidden Conflicts

After several preliminary studies, two types of hidden conflict were introduced in [8]. We speak either about *internal* hidden conflict of given BF or about a mutual hidden conflict *between* two BFs. Let us recall the hidden conflict definitions and their most important properties here. For introductory examples and more details see [5, 7, 8].

We shall note that hidden conflict and its degrees are just extensions of classic Shafer's definition of conflict. It is not a new alternative definition or approach.

Definition 1 *Assume two BFs m^i and m^{ii} such that for some $k>0$ $(\circledcap_1^k(m^i \circledcap m^{ii}))(\emptyset) = 0$. If there further holds $(\circledcap_1^{k+1}(m^i \circledcap m^{ii}))(\emptyset) > 0$ we say that there is a* conflict of BFs m^i and m^{ii} hidden in the k-th degree *(hidden conflict of k-th degree, abbreviated as HC_k). If there is already $(\circledcap_1^{k+1}(m^i \circledcap m^{ii}))(\emptyset) = (m^i \circledcap m^{ii}))(\emptyset) > 0$ for $k = 0$ then there is a conflict of respective BFs which is not hidden or we can say that it is* conflict hidden in degree zero *(HC_0).*

Theorem 1 *Hidden conflict of non-vacuous BFs on $\Omega_n, n > 1$ is always of a degree less or equal to $n-2$; i.e., the condition*

$$(\circledcap_1^{n-1}(m^i \circledcap m^{ii}))(\emptyset) = 0 \tag{1}$$

always means full non-conflictness of respective BFs and there is no hidden conflict.

Definition 2 *Let us assume a BF is given by m such that $(\circledcap_1^2 m)(\emptyset) = 0$ and $(\circledcap_1^s m)(\emptyset) > 0$ for an $s > 2$. Then we say that there is* an internal hidden conflict *in m. More specifically, if $\exists k \geq 0$ such that $(\circledcap_1^{k+1} m)(\emptyset) = 0$ and $(\circledcap_1^{k+2} m)(\emptyset) > 0$, then we say that there is* an internal conflict of BF m hidden in k-th degree[1] *– hidden internal conflict of k-th degree (HIC_k).*

Theorem 2 *Internal hidden conflict of any BF on Ω_n, $n > 1$ is always of a degree less or equal to $n-2$; i.e., the condition*

$$(\circledcap_1^n m)(\emptyset) = 0 \tag{2}$$

always means the full internal non-conflictness of any BF given by any bba m on any Ω_n.

Theorem 3 *(i) Let us assume two BFs m^i, m^{ii} with hidden conflict of k-th degree for $k \geq 2$ and their conjunctive combination $m = m^i \circledcap m^{ii}$. Then there is an internal conflict of m hidden in $k-1$-th degree.*

(ii) Any hidden conflict of any BF m of any degree $k > 1$ can be expressed as a hidden conflict of two BFs of degree $k+1$: $m = m \circledcap m_{vac}$.

Proof. For proofs of both the theorems see [8].

[1] Note that for $k = 0$ there is just $m(\emptyset) = 0$ and $(m \circledcap m)(\emptyset) > 0$; hence m is consistent and the internal conflict is not hidden or we can say *hidden in degree zero*; for $k = 1$ there is just $(m \circledcap m)(\emptyset) = 0$ and $(m \circledcap m \circledcap m)(\emptyset) > 0$ hence the conflict is hidden in the 1^{st} degree.

Note that the definition of k-th degree of hidden conflict differs in powers of $\textcircled{\cap}^k$ (see Definitions 1 and 2). The reason is straightforward: Note that while $(m_1 \textcircled{\cap} m_2)(\emptyset) = (\textcircled{\cap}^1(m_1 \textcircled{\cap} m_2))(\emptyset) > 0$ represents conflict which is not hidden, (i.e., hidden in degree 0), it is already $(\textcircled{\cap}^2(m_1 \textcircled{\cap} m_2))(\emptyset) > 0$ which represents hidden conflict in the degree 1. On the other hand, $(m \textcircled{\cap} m)(\emptyset) = (\textcircled{\cap}^2 m)(\emptyset) > 0$ represents internal conflict (of one BF) (i.e., hidden in degree 0). Thus the first degree of internal hidden conflict appears for $(\textcircled{\cap}^3 m)(\emptyset) > 0$ and k-th degree for $(\textcircled{\cap}^{k+1} m)(\emptyset) = 0$ while $(\textcircled{\cap}^{k+2} m)(\emptyset) > 0$.

Definition 3 *(i) Assume two BFs m^i and m^{ii}. We say that the BFs are non-conflicting in k-th degree if $(\textcircled{\cap}_1^k (m^i \textcircled{\cap} m^{ii})(\emptyset) = 0$.*
(ii) BFs m^i and m^{ii} are fully non-conflicting *if they are non-conflicting in any degree.*

Theorem 4 *Any two BFs on n-element frame of discernment Ω_n non-conflicting in the n-th degree are fully non-conflicting.*

Proof. For the idea of the proof see [5].

In this study, we are not interested in a numeric size of any conflict. What we are interested in are conflictness and non-conflictness. As all the degrees of hidden conflicts are only extensions of classic conjunctive conflict $(m_1 \textcircled{\cap} m_2)(\emptyset)$, all conflictness/non-conflictness depend only on *the sets of focal elements* $\mathcal{F}_1, \mathcal{F}_2$ — on the structures of respective BFs — not on their bbms. More specifically, it depends on the number and cardinalities of focal elements, their intersections, nestedness, etc. We put the masses corresponding to particular focal elements aside in our examples and focus only on the structures of sets of focal elements.

4 Extension and Correction of Hidden Conflict Theory

4.1 Hidden Conflict on Ω_2 and Hidden Conflict of $(n-1)$-th Degree

When preparing this study we have observed hidden conflict also on $\Omega_2 = \{a, b\}$ and analogously we can find a hidden conflict of $(n-1)$-th degree on any n-element frame of discernment. How it is possible? According to the previous section and namely to [6, 8] the maximal degree of hidden conflict is $n-2$ on Ω_n. Since $2-2=0$, only a conflict which is not hidden is possible on Ω_2. There is $(m_{vac} \textcircled{\cap} m)(\emptyset) = m(\emptyset) = 0$ for any normalised BF m. Nevertheless, $(\textcircled{\cap}^2(m_{vac} \textcircled{\cap} m))(\emptyset) = (m \textcircled{\cap} m)(\emptyset) > 0$ whenever $m(a) > 0, m(b) > 0$, thus for any general BF m with both the singletons. How this is possible?

The reason is the following. We were looking for a hidden mutual conflict between two BFs in [6, 8]. m_{vac} is considered to be non-conflicting with any BF. Therefore, it is non-conflicting also with any m with both singletons among its focal elements. Hence the hidden conflict of m_{vac} and m is just an internal conflict of m, see also $(m \textcircled{\cap} m)(\emptyset) > 0$ above. I.e., it is an internal conflict of m which is not hidden. In the case of combination, it is hidden by m_{vac}.

The analogous situations appear on any finite frame of discernment; hence we obtain:

Lemma 1 *Let m_1 and m_2 be two belief functions on Ω_n such that they have hidden conflict of $(n-1)$-th degree. Then $\mathcal{F}_1$ contains n subsets of Ω_n of cardinality $n-1$ only, and possibly also entire Ω_n, $\mathcal{F}_2 = \{\Omega_n\}$ or vice versa. Hence one the BFs is vacuous and the corresponding hidden conflict is in fact an internal hidden conflict the other one.*

Proof. Assertion follows Theorem 6 and the above text of this section.

Our current observation and Lemma 1 show the importance of distinguishing internal conflict and of entire/global conflict of two BFs from mutual conflict between them [2] also in a hidden case!

4.2 Belief Structures in Hidden Conflict on Ω_3

Let us present a correction of Lemma 5 from [8] about structures of non-vacuous BFs which have hidden conflict of the $(n-2)$-th degree. The original statement of the Lemma holds for all frames for $n > 3$. There are more belief structures for smaller frames Ω_2 and Ω_3. The corrected version is the following:

Lemma 2 *(i) The only non-vacuous BFs on Ω_n with hidden conflict of degree $(n-2)$ are BFs with focal elements of cardinality $\geq n-1$ for any $n > 3$, such that one has at least $(n-1)$ focal elements of cardinality $(n-1)$ and the other one has just one focal element of cardinality $(n-1)$. Moreover, every $(n-1)$-element subset of Ω_n must be a focal element of either one or both BFs.*
(ii) The only non-vacuous BFs on Ω_n with hidden conflict of degree $(n-2)$ are BFs with focal elements of cardinality $\geq n-1$ for $n = 2, 3$, such that each of them has at least one focal elements of cardinality $(n-1)$ and moreover, every $(n-1)$-element subset of Ω_n must be a focal element of either one or both BFs.

Proof. For proof and explanation see Appendix I.

5 Theoretic Approach

Let suppose a pair of BFs m_1, m_2 with focal elements $\mathcal{F}_1, \mathcal{F}_2$. If $(m_1 \odot m_2)(\emptyset) > 0$, there is a non-hidden conflict, i.e., if there exists $A \in \mathcal{F}_1, B \in \mathcal{F}_2$ with non-empty intersection $A \cap B = \emptyset$. On the other hand the simplest case of non-conflictness of the 1-st degree is characterized by $(m_1 \odot m_2)(\emptyset) = 0$, i.e., by $A \cap B \neq \emptyset$ for any $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$.

What does it mean conflict hidden in degree 1? According to main definition of hidden conflict, Definition 1, a hidden conflict of the first degree arises whenever

$$(m_1 \odot m_2)(\emptyset) = 0 \text{ and } ((m_1 \odot m_2) \odot (m_1 \odot m_2))(\emptyset) > 0.$$

Hence we can characterise the class of pairs of BFs non-conflicting in the second degree by $(\bigodot^2(m_1 \odot m_2))(\emptyset) = 0$. Let us turn our attention to the motivation of HC, conflict observation, and the original working definition of hidden conflict: There is the principal assumption that the combination $(m_1 \odot m_2)$ of two mutually non-conflicting BFs m_1, m_2 should be non-conflicting with any of the original m_1 and m_2, hence both

$(m_1 \circledcirc (m_1 \circledcirc m_2))(\emptyset) = 0$ and $((m_1 \circledcirc m_2) \circledcirc m_2)(\emptyset) = 0$. Using associativity and commutativity of conjunctive combination $\circledcirc$ we have $((m_1 \circledcirc m_2) \circledcirc (m_1 \circledcirc m_2))(\emptyset) = ((m_1 \circledcirc m_1) \circledcirc (m_2 \circledcirc m_2))(\emptyset)$ and it is zero whenever any focal element of $(m_1 \circledcirc m_1)$ has non-empty intersection with any focal element of $(m_2 \circledcirc m_2)$ and vice versa.

Thus the class of non-conflictness of the 2-nd degree is specified by $(\bigcirc^2 (m_1 \circledcirc m_2))(\emptyset) = 0$ and alternatively by $(X_i \cap X_j) \cap (Y_r \cap Y_s) \neq \emptyset$ for any focal elements $X_i, X_j \in \mathcal{F}_1$, $Y_r, Y_s \in \mathcal{F}_2$ of m_1, m_2. Note that the first condition corresponds to Yager's pair-wise consistency of $m_{12} = (m_1 \circledcirc m_2)$, which appears if $\sum_{X \cap Y \neq \emptyset, X, Y \in \mathcal{F}_{12}} m_{12}(X) m_{12}(Y) = 1$, see [16].

Analogously, we can continue to hidden conflicts and classes of non-conflictness of higher degrees: hidden conflict of the 2-nd degree and non-conflictness of the 3-rd degree, up to hidden conflict of the $(k-1)$-th degree and related non-conflictness of the k-th degree. Analogously to the classes of the 1-st and 2-nd degrees, we have also two characterizations of the class: (i) one based on the original bbas m_1 and m_2: $\bigcap_1^k X_i \cap \bigcap_1^k Y_j \neq \emptyset$ for any k-tuples of focal elements $X_i \in \mathcal{F}_1$ and $Y_j \in \mathcal{F}_2$, and (ii) the other characterization based on combination $m_{12} = m_1 \circledcirc m_2$: $(\bigcirc^k (m_1 \circledcirc m_2))(\emptyset) = 0$.

Ad (i): If either m_1 or m_2 has less focal elements than k, the focal elements are repeating in the computation of hidden conflict; see, e.g., one of the m_i's in the Introductory and the Little Angel Examples, see [8], thus analogously also in the verification of non-conflictness. Hence intersecting X_i, Y_i need not be different. Hence intersection of any k-tuple of elements of $\mathcal{F}_1$ (possibly with repeating) must be non-empty and must have a non-empty intersection with the intersection of any k-tuple of elements of $\mathcal{F}_2$ (possibly with repeating).

Ad (ii): this correspond to Pichon et al.'s k-consistency of m_{12}, see [12]; for $k = n$ to logical consistency [9].

Theorem 5 *For any pair of BFs Bel_1, Bel_2 given by m_1, m_2 the following is equivalent:*
(i) Bel_1 and Bel_2 are non-conflicting in degree k.
(ii) $\bigcap_1^k X_i \cap \bigcap_1^k Y_j \neq \emptyset$ for any k-tuples of focal elements $X_i \in \mathcal{F}_1$, $Y_j \in \mathcal{F}_2$ of m_1, m_2.
(iii) $(\bigcirc^k (m_1 \circledcirc m_2))(\emptyset) = 0$ for $m_{12} = m_1 \circledcirc m_2$.

6 Descriptive Approach

Let us start with Ω_2. There are $2^3 - 1 = 7$ different belief function structures. We can create a 7×7 table of all pairs of these structures as in Table 1. The black dot represents a singleton, and the black oval is the focal element of cardinality 2, which corresponds to Ω_2 in this case. Table cells correspond to $\circledcirc$ combination of respective structures. Since the $\circledcirc$ operator is commutative, only the right upper part is filled in. White cells correspond to non-conflicting structures, red and cyan to conflicting ones (red represents total conflict). The 4 green cells correspond to hidden conflict, as described in Lemma 1.

This case of Ω_2 has an excellent interpretation. We can easily see that non-conflicting pairs are just the consonant ones (including m_{vac}). Note that either one of them is m_{vac} or both contain the same singleton — the white cells of the table. Conflicting pairs (HC_0)

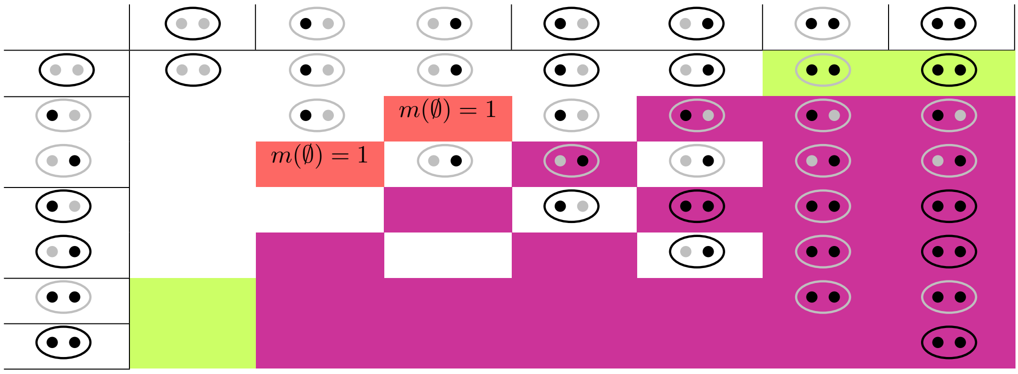


Table 1: All possible combinations of belief function structures on Ω_2

are of two types: (i) one structure contains both singletons and the other is non-vacuous. (ii) both structures have just one singleton, each different. And finally, 4 green HC_1 fields correspond to Lemma 1

Let us continue on Ω_3. This case is significantly more complex. There are $2^7 - 1 = 127$ belief structures here, (note that we have $2^{2^n-1} - 1$ possible structures in general), To give the reader a similar impression as from Table 1, we created a 127×127 bitmap — see Figure 1. Similarly to Ω_2 rows and columns correspond to structures. The structures are ordered by the number of focal elements as the first criterion and their size as the second one. I.e. the ordering is the following: $\{a\}; \{b\}; \{c\}; \{ab\}; \{ac\}; \{bc\}; \{abc\}; \{a, b\}, \{a, c\}, \ldots$. Therefore, e.g. the 7th row and column correspond to vacuous BF. White cells correspond to non-conflict situations, red to HC_0, orange to HC_1, and black to HC_2. Striped cells in Figure 1(a) correspond to pure type of respective conflict as defined later. Note that black cells corresponding to HC_2 appears in row and column corresponding to vacuous BF only.

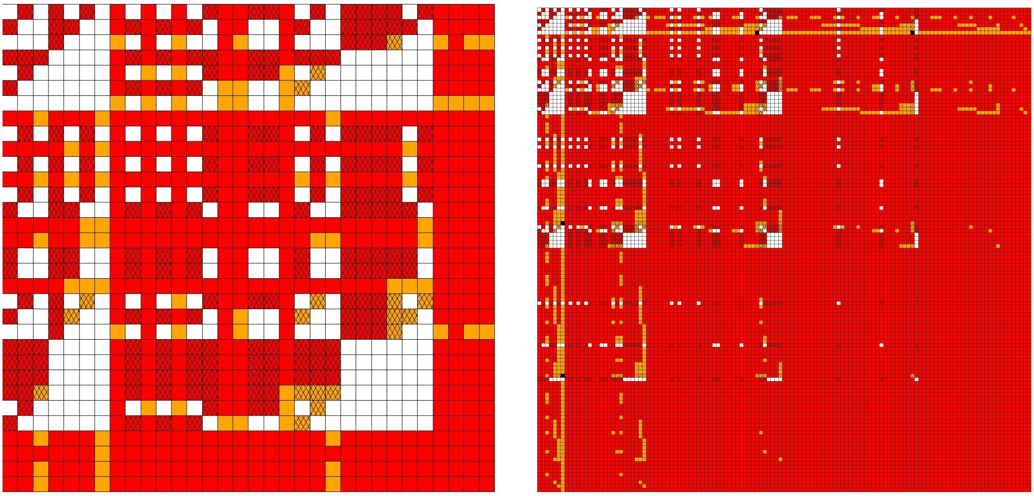
We can easily see white, i.e., non-conflicting area at (23–28)x(23–28) and other areas (23–28)x(60-63), (60-63)x(23-28), and (60-63)x(60-63). Where 22–26 are two couples, 27–28 couple and triple, 60–63 structures with 3 focal elements all 22–28, 60–63 contains c. Thus this is not theoretically very interesting; this area comes from the selected ordering of belief structures. The complete analysis of the bitmap is still under preparation.

Nevertheless, the bitmaps are already part of the experimental results, thus related to the next section.

7 Experimental Approach

Conflictness/non-conflictness according to the classical definition of (conjunctive) conflict depends only on the structure of the focal elements given by the bbas. In this section, we will show some results of experiments with these structures.

A conflict based on a mass assigned to an empty set by the conjunctive rule has two



(a) Zoom of left upper part (32 × 32) (b) Full 127 × 127 bitmap

Figure 1: Bitmap of combination of structures on Ω_3

levels. The first relates to the very existence of the conflict, i.e., whether there are two focal elements with an empty intersection. The second level deals with the magnitude of the conflict. The size corresponds to the number of pairs of focal elements with empty intersections and the probability masses they carry.

This paper focuses on the first part of the problem - the theoretical possibility of conflict which is connected with the structure of focal elements only. We are not interested in the probability mass assigned to individual focal elements here.

We know that the number of different structures on a given frame of discernment is super-exponential with respect to frame size. See the first column in Table 2. Suppose we disregard the frame of discernment labelling. Then we can group structures into permutation-equivalent classes and calculate individual properties for only one representative of each class. We can say that we are creating a certain catalogue of structures of belief functions. The number of classes of permutation-equivalent structures for $\Omega_2, \Omega_3, \Omega_4$, and Ω_5 is in Table 2. Note that we were not able to create this catalogue for frame of discernment having more than five elements. Please, be aware that the exact number of classes of permutation equivalent structures for Ω_5 is unknown. By the submission deadline, we were not able to go through all the structures with 15 and 16 focal elements. Therefore, the number from Table 2 is an estimate based on the number of classes for structures with other numbers of focal elements.

How do you recognize that two structures are permutation equivalent? It turns out that the problem corresponds to graph isomorphism – bipartite graphs isomorphism specifically. In this case, focal elements are vertices on one side, and the frame of discernment is represented by vertices on the other side of the graph. To solve graph isomorphism problem, we used the BLISS algorithm by Junttila and Kaski [10, 11] as implemented in

	number of structures	classes of permutation equivalent structures
Ω_2	7	5
Ω_3	127	39
Ω_4	32.767	1.990
Ω_5	2.147.483.648	8.820*

Table 2: Richness of structures

the igraph [1] R package. Note that the algorithm is based on a special heuristic of finding the canonical form of a graph unique for isomorphism.

7.1 Equivalence Classes of Belief Functions Structures

Let us present some interesting statistics on the structures of belief functions and respective equivalence classes. Note that we plan to create an online catalogue of all classes but at the time of writing this paper it is still an ongoing process.
Ω_3: Equivalence classes have three cardinalities (1, 3, and 6). There are 7 classes with only one structure. 24 classes contain 3 structures, and 8 classes contain 6 structures.
cardinality 1: e.g., $\{abc\}$, $\{a, b, c\}$[2] or m_1 and m_2 from Example 1
cardinality 3: e.g., $\{a\}$, $\{a, ab, ac\}$, $\{a, bc, abc\}$
cardinality 6: e.g., $\{a, ab\}$, $\{a, ab, abc\}$, $\{b, ab, ac\}$

Ω_4: In the case of Ω_4, equivalence classes have seven cardinalities (1, 3, 4, 6, 12, 24, and 48). There are 15 classes with only one structure. 16 classes contain 3 structures, and 894 classes contain 24 structures. Interestingly, there is only one class with 48 structures. One of these 48 structures is $\{a, ab, bc, bcd\}$. Note that this structure has internal conflict.

There are more greater classes with increasing n, even for analogous structures, look at the above structures from Ω_3 on greater frame Ω_5:
cardinality 1: e.g., $\{abcde\}$, $\{a, b, c, d, e\}$
cardinality 5: e.g., $\{a\}$,
cardinality 20: e.g., $\{a, ab\}$
cardinality 50: e.g., $\{a, ab, ac\}$
cardinality 60: e.g., $\{a, ab, abc\}$
cardinality 100: e.g., $\{b, ab, ac\}$

7.2 Internal Conflict

A proper survey of all structures aims to provide a detailed insight into the internal structure of individual conflicts and their types. In the case of one belief function, we recognize a hidden internal conflict of the structure of various degrees.

By internal conflict we mean the conflict which is inside a single bba, caused by conflicting masses of the bba, it may appear when we combine the bba with itself or it may remain hidden in some degree, see [8]. For each bba of the permutation-equivalent classes, we calculated whether it is internally non-conflicting or whether it has a hidden internal conflict and of which degree. Recall that the maximum degree of hidden conflict is $n - 1$. The degree of conflict hiddeness is k if $\textcircled{\cap}^{k+1} m(\emptyset) = 0$ while $\textcircled{\cap}^{k+2} m(\emptyset) > 0$.

[2]We use abbreviations $\{abc\}$ for $\{\{abc\}\}$, $\{a, b, c\}$ for $\{\{a\}, \{b\}, \{c\}\}$ and analogous, in this section.

In the following table we use the following notation: NC — non-conflicting, IC — internal conflict, HIC_k - hidden internal conflict of k-th degree. The numbers in parentheses refer to all structure. Other numbers to classes of equivalence.

	NC	IC	HIC_1	HIC_2
Ω_2	3(5)	2(2)	–	–
Ω_3	11(37)	26(88)	2(2)	–
Ω_4	79(941)	1.867(31.392)	42(432)	2(2)

Table 3: Number of classes with different internal conflictness/non-conflicness

Example 1 *The only two classes of belief function structures allowing maximal hidden internal conflict for each frame of discernment are represented by the following structures $\mathcal{F}_1, \mathcal{F}_2$ for various frames of discernment:*

- $\Omega_2 = \{a, b\}$*:* $\mathcal{F}_1 = \{a, b\}$*;* $\mathcal{F}_2 = \{a, b, ab\}$
- $\Omega_3 = \{a, b, c\}$*:* $\mathcal{F}_1 = \{ab, ac, bc\}$*;* $\mathcal{F}_2 = \{ab, ac, bc, abc\}$
- $\Omega_4 = \{a, b, c, d\}$*:* $\mathcal{F}_1 = \{abc, abd, acd, bcd\}$*;* $\mathcal{F}_2 = \{abc, abd, acd, bcd, abcd\}$

Generally for Ω_n, the focal elements have to be all subsets of cardinality $|n-1|$ with possible focal element covering the whole Ω_n. $\mathcal{F}_1 = \{A \subset \Omega_n : |A| = n-1\}$;$\mathcal{F}_2 = \{A : A \in \mathcal{F}_1$ or $A = \Omega_n\}$. Then, the respective BFs have internal conflict hidden in $(n-1)$-th degree. This corresponds to Theorem 15 in [8].

7.3 Mutual Conflict

In case of two different bbas, their mutual conflict can be also hidden. Assume m_1 and m_2. The definition of hidden conflict of k degree is that $(\textcircled{\cap}^k(m_1 \textcircled{\cap} m_2))(\emptyset) = 0$ while $(\textcircled{\cap}^{k+1}(m_1 \textcircled{\cap} m_2))(\emptyset) > 0$. In this experiment we tried to distinguish mutual conflict from false mutual conflict caused by possible (hidden) internal conflict of one of the involved bbas. To enumerate all pairs we employ the fact that we already have a catalogue of permutation equivalent structures. Technically, instead of going through all possible pairs of structures, we used only representatives from each permutation equivalent class on the one hand. On the other hand, we had to go through all the structures. This explains why we do not have results for Ω_5. The total number of pairs with a given property is then obtained by multiplying the sizes of a given permutation-equivalent class of structures. The symmetry of the whole operation guarantees the correct result.

Assume $m = m_1 \textcircled{\cap} m_2$. In the following table we use also this notation:

P — pure mutual hidden conflict: $(\textcircled{\cap}^k m)(\emptyset) > 0$ and $(\textcircled{\cap}^n m_1)(\emptyset) = 0$, $(\textcircled{\cap}^n m_2)(\emptyset) = 0$,

C — clear degree of mutual hidden conflict: i.e., $(\textcircled{\cap}^k m)(\emptyset) > 0$ and $(\textcircled{\cap}^k m_1)(\emptyset) = 0$, $(\textcircled{\cap}^k m_2)(\emptyset) = 0$ (degree comes from mutual, not from internal conflict(s)),

F — hidden conflict which may to be caused by internal conflict of either m_1 or m_2. $(\textcircled{\cap}^k m)(\emptyset) > 0$ and simultaneously $(\textcircled{\cap}^k m_1)(\emptyset) > 0$ or $(\textcircled{\cap}^k m_2)(\emptyset) > 0$;

unfortunately we cannon distinguish whether it is false mutual hidden conflict or a mixture of mutual and internal conflicts in general.

	NC	HC_0		HC_1			HC_2			HC_3		
		P=C	F	P	C	F	P	C	F	P	C	F
Ω_2	17	8	20	0	0	4						
Ω_3	649	672	14.048	48	100	656	0	0	4			
Ω_4	258.785	582.016	1.071.094.400	46.696	283.708	1.454.784	32	64	2.528	0	0	4

We have to note that, surprisingly less numbers of HC_2 (both P and C; 32 and 64) on Ω_4 than HC_1 (both P and C; 48 and 100) on Ω_3 (both the cases are conflict hidden on $(n-1)$-th degree on corresponding Ω_n) comes from the situation described in Lemma 1. There are more corresponding structures on Ω_3 than on Ω_4.

Example 2 *(i) Note that the 4 pairs of maximum hidden degree corresponds to both m_1, m_2 from Example 1; they are all combinations of m_1, m_2 with vacuous bba m_{vac}: $m_1 \circledcirc m_{vac}$, $m_2 \circledcirc m_{vac}$, $m_{vac} \circledcirc m_1$, and $m_{vac} \circledcirc m_2$ on any frame. It is generally assumed, that m_{vac} is mutually non-conflicting with any other bba, hence conflicts with the maximum degree of hiddeness $n-2$, are always false, they are always internal hidden conflicts of one of the bbas.*
(ii) Analogously, two numerically identical bbas $m_i \equiv m_j$ with an internal hidden conflict are mutually non-conflicting, hence internal hidden conflict of $m_i \circledcirc m_j$ is also false. Specially, also $m_i \circledcirc m_i$ for bbas from Example 1, nevertheless this time of less degree of hiddeness (degree $\lceil n-2/2 \rceil$).

8 Conclusion

Several theoretic extensions and corrections related to maximal degree of hidden conflict have been presented. Theoretic chracterization of classes of non-conflictness of belief functions has been formulated.

Descriptive and experimental approaches to analysis of combination of belief function structures have been presented. A catalogue of structures of belief functions and of combination of these structures is under preparation.

Acknowledgement

This work was supported by the institutional support RVO: 67985807 (first author) and grant GA ČR 19-04579S (second author)

References

[1] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[2] Milan Daniel. Conflicts within and between Belief Functions. In Eyke Hüllermeier, Rudolf Kruse, and Frank Hoffmann, editors, *IPMU 2010*, volume 6178 of *LNCS*, pages 696–705. Springer, Berlin, Heidelberg, 2010.

[3] Milan Daniel. Non-conflicting and Conflicting Parts of Belief Functions. In *7th International symposium on Imprecise Probability: Theories and Applications (ISIPTA 2011)*, pages 149–158, Innsbruck, 2011. SIPTA.

[4] Milan Daniel. Conflict between Belief Functions: A New Measure Based on Their Non-conflicting Parts. In F. Cuzzolin, editor, *BELIEF 2014*, volume 8764 of *LNCS*, pages 321–330, Cham, 2014. Springer.

[5] Milan Daniel and Václav Kratochvíl. Belief Functions and Degrees of Non-conflictness. In G. Kern-Isberner and Z. Ognjanović, editors, *ECSQARU 2019*, volume 11726 of *LNCS*, pages 125–136, Cham, 2019. Springer.

[6] Milan Daniel and Václav Kratochvíl. On Hidden Conflicts of Belief Functions. In *11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2019)*. Atlantis Press, 2019.

[7] Milan Daniel and Václav Kratochvíl. Comparison of shades and hiddenness of conflict. In J. Vejnarová and N. Wilson, editors, *ECSQARU 2021*, volume 12897 of *LNCS*, pages 314–327, Cham, 2021. Springer.

[8] Milan Daniel and Václav Kratochvíl. Hidden conflicts of belief functions. *International Journal of Computational Intelligence Systems*, 14:438–452, 2020.

[9] Sébastien Destercke and Thomas Burger. Toward an Axiomatic Definition of Conflict Between Belief Functions. *IEEE transactions on cybernetics*, 43(2):585–596, 2013.

[10] Tommi Junttila and Petteri Kaski. Engineering an efficient canonical labeling tool for large and sparse graphs. In *2007 Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 135–149. SIAM, 2007.

[11] Tommi Junttila and Petteri Kaski. Conflict propagation and component recursion for canonical labeling. In A. Marchetti-Spaccamela and M. Segal, editors, *TAPAS 2011*, volume 6595 of *Lecture Notes in Computer Sci.*, pages 151–162. Springer, 2011.

[12] Frédéric Pichon, Anne-Laure Jousselme, and Nadia Ben Abdallah. Several shades of conflict. *Fuzzy Sets and Systems*, 366:63–84, 2019.

[13] Glenn Shafer. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.

[14] Philippe Smets. Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2):133–147, 2005.

[15] Milan Studený and Václav Kratochvíl. Facets of the cone of exact games. *Mathematical Methods of Operations Research*, 95(1):35–80, 2022.

[16] Ronald R. Yager. On Considerations of Credibility of Evidence. *International Journal of Approximate Reasoning*, 7:45–72, 1992.

Appendix I. Theoretical Corrections — Extensions

Let us present the original version of the theorem on maximal degree of hidden conflict (Theorem 2 in [8]) and a generalisation of the theorem and its corollary.

Theorem 6 (maximum degree of hidden conflict; original version) *For any non-vacuous BFs Bel^i, Bel^{ii} defined by m^i and m^{ii} on Ω_n it holds that*

$$(\textcircled{\cap}_1^{n-1}(m^i \textcircled{\cap} m^{ii}))(\emptyset) = 0 \quad \textit{iff} \quad (\textcircled{\cap}_1^{k}(m^i \textcircled{\cap} m^{ii}))(\emptyset) = 0$$

for any $k \geq n-1$.

Corollary 1 (original) *A hidden conflict of any non-vacuous BFs on any Ω_n always has has degree less than or equal to $n-2$; i.e., the condition*

$$(\textcircled{\cap}_1^{n-1}(m^i \textcircled{\cap} m^{ii}))(\emptyset) = 0 \tag{3}$$

always means the full non-conflictness of any BFs m^i and m^{ii} on any Ω_n. Moreover, there is no hidden conflict on any two-element frame Ω_2

The original version of the theorem is O.K. in [8] because non-vacuous BFs are expectected there. Nevertheless, we can generalized its assertion as it follow:

Theorem 7 (maximum degree of hidden conflict; gneralized) *(i) For any BFs Bel^i, Bel^{ii} defined by m^i and m^{ii} on Ω_n it holds that*

$$(\textcircled{\cap}_1^{n}(m^i \textcircled{\cap} m^{ii}))(\emptyset) = 0 \quad \textit{iff} \quad (\textcircled{\cap}_1^{k}(m^i \textcircled{\cap} m^{ii}))(\emptyset) = 0$$

for any $k \geq n$.
(ii) For any non-vacouous BFs Bel^i, Bel^{ii} the stronger assertion holds true for any $k \geq n-1$

$$(\textcircled{\cap}_1^{n-1}(m^i \textcircled{\cap} m^{ii}))(\emptyset) = 0 \quad \textit{iff} \quad (\textcircled{\cap}_1^{k}(m^i \textcircled{\cap} m^{ii}))(\emptyset) = 0.$$

Proof. The assertion follow the proof in [8] and the text in Section 4.1.

The original version of the corollary need a small correction — specfication of mutual conflict between BFs — see the second assertion of the generalised version:

Corollary 2 (generalised) *(i) A hidden conflict of any two BFs on any Ω_n always has a degree less than or equal to $n-1$; i.e., the condition*

$$(\textcircled{\cap}_1^{n}(m^i \textcircled{\cap} m^{ii}))(\emptyset) = 0 \tag{4}$$

always means the full non-conflictness of any two BFs m^i and m^{ii} on any Ω_n.
(ii) A hidden conflict of any non-vacuous BFs on any Ω_n always has a degree less than or equal to $n-2$; i.e., the condition

$$(\textcircled{\cap}_1^{n-1}(m^i \textcircled{\cap} m^{ii}))(\emptyset) = 0 \tag{5}$$

always means the full mutual non-conflictness between any two BFs m^i and m^{ii} on any Ω_n. Specially, there is neither a hidden mutual conflict nor a hidden internal conflict[3] *between any two BFs on two-element frame Ω_2.*

[3]There may to be only mutual conflicts of degree $2-2=0$; meaning there are only mutual conflicts $(m^i \textcircled{\cap} m^{ii})(\emptyset) > 0$ and internal conflicts $(m \textcircled{\cap} m)(\emptyset) > 0$ which are not hidden.

Let us present a compariosn of the original and updated versions of Lemma 5 in [8] now.

Lemma 3 (original version of Lemma 5) *The only non-vacuous BFs on Ω_n with hidden conflict of degree $(n-2)$ are BFs with focal elements of cardinality $\geq n-1$, such that one has at least $(n-1)$ focal elements of cardinality $(n-1)$ and the other one has just one focal element of cardinality $(n-1)$. Moreover, every $(n-1)$-element subset of Ω_n must be a focal element of either one or both BFs.*

Lemma 4 (updated version of Lemma 5) *(i) The only non-vacuous BFs on Ω_n with hidden conflict of degree $(n-2)$ are BFs with focal elements of cardinality $\geq n-1$ for any $n > 3$, such that one has at least $(n-1)$ focal elements of cardinality $(n-1)$ and the other one has just one focal element of cardinality $(n-1)$. Moreover, every $(n-1)$-element subset of Ω_n must be a focal element of either one or both BFs.*
(ii) The only non-vacuous BFs on Ω_n with hidden conflict of degree $(n-2)$ are BFs with focal elements of cardinality $\geq n-1$ for $n = 2, 3$, such that each of them has at least one focal elements of cardinality $(n-1)$ and moreover, every $(n-1)$-element subset of Ω_n must be a focal element of either one or both BFs.

Proof. The characterisation of the 'other' BF is based on the fact that addition of any other focal element of cardinality $n-1$ decreases focal elements by $\circledcirc$, hence also decreases a degree of conflict hiddeness.

This is true in general. Nevertheless this is not a matter on Ω_3: as $(m_1 \circledcirc m_1) \circledcirc (m_2 \circledcirc m_2)$ cannot to decrease focal element twice; yes, there are three operations $\circledcirc$, each of them theoretically may to decrease the size of focal elements, but $n-1 = 2$ is decreased to zero already by two operations $\circledcirc$, hence other combination cannot further decrease the size of focal elements (decrease of cardinality of empty set). Hence both m_i may contaning two or three focal elements of cardinality 2 and highest degree 1 of hidden conflict is kept. Similarly, $n-1 = 1$, thus size of focal elements and (zero) degree of hidden conflict cannot be decreased twice, even if both singletons are in both bbas m_1 and m_2.

Computing the Decomposable Entropy of Graphical Belief Function Models

Radim Jiroušek[1,2], Václav Kratochvíl[1,2], and Prakash P. Shenoy[3]

[1]Faculty of Management, University of Economics, Jindřichův Hradec, Czechia
[2]Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czechia
[2] *radim@utia.cas.cz, velorex@utia.cas.cz*
[3]School of Business, University of Kansas, Lawrence, KS, USA
[3] *pshenoy@ku.edu*

June 13, 2022

Abstract

In 2018, Jiroušek and Shenoy proposed a definition of entropy for Dempster-Shafer (D-S) belief functions called *decomposable entropy*. Here, we provide an algorithm for computing the decomposable entropy of directed graphical D-S belief function models. For undirected graphical belief function models, assuming that each belief function in the model is non-informative to the others, no algorithm is necessary. We compute the entropy of each belief function and add them together to get the decomposable entropy of the model. Finally, the decomposable entropy generalizes Shannon's entropy not only for the probability of a single random variable but also for multinomial distributions expressed as directed acyclic graphical models called Bayesian networks.

1 Introduction

Jiroušek and Shenoy (2018a) propose a definition of entropy for Dempster-Shafer (D-S) belief functions called *decomposable entropy*. Some basic properties of the decomposable entropy are described in (Jiroušek and Shenoy, 2020). One of the main properties of this entropy is as follows. Suppose we have a joint basic probability assignment (BPA) $m_{X,Y}$ for $\{X, Y\}$ that decomposes as follows: $m_{X,Y} = m_X \oplus m_{Y|X}$, where m_X is the marginal of $m_{X,Y}$ for X, $m_{Y|X}$ is a conditional BPA for Y given X, and $\oplus$ is Dempster's combination rule. Then, the joint decomposable entropy of $m_{X,Y}$, denoted by $H(m_{X,Y})$, is equal to $H(m_X) + H(m_{Y|X})$, where $H(m_{Y|X})$ denotes the conditional decomposable entropy of $m_{Y|X}$. This decomposable property is analogous to the decomposable property of Shannon's entropy for joint probability mass functions that is the basis of its definition (Shannon, 1948). There are numerous definitions of entropy for the D-S theory (see (Jiroušek and Shenoy, 2018b) for a review), but none of these satisfy the decomposable property and, therefore, the computation of these entropies for large graphical models may be intractable.

Graphical belief function models can be either directed or undirected. This article provides an algorithm for computing the decomposable entropy of directed graphical belief function models and illustrates it using an example called the *captain's decision problem* (Almond, 1995). This problem has eight variables, and the joint state space of the eight variables has 2,304 states.

Two distinct belief functions are said to be *non-informative* if the marginals of these belief functions for the intersection of their domains are vacuous. A set of distinct belief functions is said to be non-informative if every pair of belief functions from the set is non-informative. No algorithm is necessary for undirected graphical belief function models with non-informative belief functions. We compute the entropy of each belief function in the model and add them together to get the entropy of the model. This is illustrated by using the *communication network* example (Haenni and Lehmann, 2002). This problem has forty-six binary variables with a joint state space of 2^{46} states and seventy non-informative belief functions.

Finally, the decomposable entropy generalizes Shannon's entropy for the probability of large multinomial distributions expressed as directed acyclic graph models called Bayesian networks. We illustrate this using the *chest clinic* Bayesian network example (Lauritzen and Spiegelhalter, 1988). First, we convert all probability potentials in the example to belief functions. In particular, we use Smets' conditional embedding to convert the conditional probability tables (CPTs) to conditional belief functions. These conditional belief functions are not Bayesian. Next, we compute the decomposable entropy of the directed graphical belief function model and show that it is the same as Shannon's entropy of this probability model. This example has eight binary variables with a joint state space of $2^8 = 256$ states.

2 Dempster-Shafer's Belief Function Theory

In this section, we sketch the basics of Dempster-Shafer‘s theory of belief functions (Dempster, 1968; Shafer, 1976).

2.1 Representations

There are several representations in the D-S theory of belief functions. Here we focus on basic probability assignments and commonality functions.

Basic Probability Assignment Suppose X is a random variable with a finite state space Ω_X. Let 2^{Ω_X} denote the set of all subsets of Ω_X. A basic probability assignment

(BPA) m for X is a function $m : 2^{\Omega_X} \to [0, 1]$ such that:

$$m(\emptyset) = 0, \text{and} \quad (1)$$

$$\sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) = 1. \quad (2)$$

$m(\mathsf{a})$ represents the probability mass that is assigned exactly to subset a. Thus, no mass is assigned to the empty subset (Eq. (1)) and the total probability assigned to all non-empty subsets is 1 (Eq. (2)).

The non-empty subsets $\mathsf{a} \in 2^{\Omega_X}$ such that $m(\mathsf{a}) > 0$ are called *focal* elements of m. A BPA m with only one focal element a (with mass 1) is called *determinsitic*. A deterministic BPA with focal element Ω_X is called *vacuous*. We say m is *consonant* if the focal elements of m are nested, i.e., if they can be ordered such that $\mathsf{a}_1 \subset \mathsf{a}_2 \subset ... \subset \mathsf{a}_m$, where $\{\mathsf{a}_1, ..., \mathsf{a}_m\}$ denotes the set of all focal elements of m. Deterministic BPAs are trivially consonant. We say m is *quasi-consonant* if the intersection of all focal elements of m is non-empty. A consonant BPA is also quasi-consonant, but not vice-versa. We say m is *Bayesian* if its focal elements are singleton subsets.

Commonality Function The information in a BPA m for X can also be represented by a corresponding commonality function (CF) Q_m for X that is defined as follows:

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} : \mathsf{b} \supseteq \mathsf{a}} m(\mathsf{b}), \text{for all } \mathsf{a} \in 2^{\Omega_X}. \quad (3)$$

$Q_m(\mathsf{a})$ represents the probability mass that could possibly move to subset a.

From Eq. (3), it follows that $0 \leq Q_m \leq 1$. From Eqs. (1)–(3), it follows that $Q_m(\emptyset) = 1$. If m is a vacuous BPA for X, then $Q_m(\mathsf{a}) = 1$ for all $\mathsf{a} \in 2^{\Omega_X}$. CFs are non-increasing in the sense that if $\mathsf{a} \subseteq \mathsf{b}$, then $Q_m(\mathsf{a}) \geq Q_m(\mathsf{b})$. The CF Q_m has exactly the same information as in the corresponding BPA m.

2.2 Marginalization and Combination

In the D-S theory, we reduce the domain of a joint belief function using the marginalization operation, and we combine distinct (or independent) belief functions using Dempster's combination rule (Dempster, 1968).

Marginalization Marginalization in D-S theory is the summation of values of BPAs.

Projection of states means dropping extra coordinates; for example, if (x, y) is a state of (X, Y), then the projection of (x, y) to X, denoted by $(x, y)^{\downarrow X}$, is simply x, which is a state of X.

Projection of subsets of states is achieved by projecting every state in the subset. Suppose $\mathsf{b} \in 2^{\Omega_{(X,Y)}}$. Then $\mathsf{b}^{\downarrow X} = \{x \in \Omega_X : (x, y) \in \mathsf{b}\}$. Notice that $\mathsf{b}^{\downarrow X} \in 2^{\Omega_X}$.

Suppose m is a BPA for (X, Y). Then, the marginal of m for X, denoted by $m^{\downarrow X}$, is a BPA for X such that for each $\mathsf{a} \in 2^{\Omega_X}$,

$$m^{\downarrow X}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_{(X,Y)}} : \mathsf{b}^{\downarrow X} = \mathsf{a}} m(\mathsf{b}). \tag{4}$$

It follows from Eq. (4), that if $m(\mathsf{b}) > 0$, then $m^{\downarrow X}(\mathsf{b}^{\downarrow X}) > 0$, for all $\mathsf{b} \in 2^{\Omega_{(X,Y)}}$.

Dempster's Combination Rule We will define Dempster's combination rule in terms of CFs. Suppose $\mathcal{X}_1$ and $\mathcal{X}_2$ are arbitrary (finite) sets of variables, and Q_1 and Q_2 are distinct CFs for $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Then $Q_1 \oplus Q_2$ is a CF for $\mathcal{X}_1 \cup \mathcal{X}_2 = \mathcal{X}$ given by:

$$(Q_1 \oplus Q_2)(\mathsf{a}) = \begin{cases} 1 & \text{if } \mathsf{a} = \emptyset, \\ K^{-1} Q_1(\mathsf{a}^{\downarrow \mathcal{X}_1})\, Q_2(\mathsf{a}^{\downarrow \mathcal{X}_2}) & \text{otherwise,} \end{cases} \tag{5}$$

for all $\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}$, where K is a normalization constant given by:

$$K = \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_{\mathcal{X}_1 \cup \mathcal{X}_2}}} (-1)^{|\mathsf{a}|+1} Q_1(\mathsf{a}^{\downarrow \mathcal{X}_1})\, Q_2(\mathsf{a}^{\downarrow \mathcal{X}_2}). \tag{6}$$

$(1 - K)$, where K is the normalization constant in Eq. (6), can be interpreted as a measure of conflict in the two CFs. The definition of Dempster's rule assumes that the normalization constant K is non-zero. If $K = 0$, i.e., $1 - K = 1$, then the two CFs Q_1 and Q_2 are said to be in *total conflict* and cannot be combined. If $K = 1$, i.e., $1 - K = 0$, we say Q_1 and Q_2 are *non-conflicting.*

Non-informative Belief Functions Suppose m_1 and m_2 are two distinct BPAs for $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. We say m_1 and m_2 are *non-informative* to each other if $m_1^{\downarrow \mathcal{X}_1 \cap \mathcal{X}_2}$ and $m_2^{\downarrow \mathcal{X}_1 \cap \mathcal{X}_2}$ are vacuous BPAs for $\mathcal{X}_1 \cap \mathcal{X}_2$. Notice that if m_1 and m_2 are non-informative to each other, then $(m_1 \oplus m_2)^{\downarrow \mathcal{X}_1} = m_1$ and $(m_1 \oplus m_2)^{\downarrow \mathcal{X}_2} = m_2$. This follows from the definition of non-informative belief functions and the local computation property (Shenoy and Shafer, 1990).

Intuitively, Q_1 doesn't tell us anything about Q_2 and vice-versa. If $\mathcal{X}_1$ and $\mathcal{X}_2$ are disjoint, then they are trivially non-informative to each other. The definition of non-informative belief functions can be generalized to sets of belief functions. A set of belief functions is non-informative if every pair of belief functions from the set is non-informative to each other. Of course, it is sufficient to check only those pairs with a non-empty intersection of their domains.

2.3 Conditional Belief Functions

Conditional belief functions were initially studied by Smets (1978) who introduced the notion of conditional embedding. They have been further explored in (Shafer, 1982; Almond, 1995; Xu and Smets, 1996). Here we review the basics.

Consider a BPA m for X and $\mathsf{a} \in 2^{\Omega_X}$. Suppose that there is a BPA for Y expressing our belief about Y if we know that $X \in \mathsf{a}$, and denote it by $m_{Y|\mathsf{a}}$. Notice that $m_{Y|\mathsf{a}} : 2^{\Omega_Y} \to [0,1]$ is a BPA for Y. We can embed this BPA for Y into a conditional BPA for (X,Y), which is denoted by $m_{\mathsf{a},Y}$, so that the following two conditions hold:

1. $m_{\mathsf{a},Y}$ tells us nothing about X, i.e., $m_{\mathsf{a},Y}^{\downarrow X}(\Omega_X) = 1$.
2. If we combine $m_{\mathsf{a},Y}$ with the deterministic BPA $m_{X\in\mathsf{a}}$ for X such that $m_{X\in\mathsf{a}}(\mathsf{a}) = 1$ using Dempster's rule, and marginalize the result to Y we obtain $m_{Y|\mathsf{a}}$, i.e., $(m_{\mathsf{a},Y} \oplus m_{X\in\mathsf{a}})^{\downarrow Y} = m_{Y|\mathsf{a}}$.

One way to obtain such an embedding is suggested by Smets (1978) (see also, (Shafer, 1982), Xu and Smets (1996), and Almond (1995)), called *conditional embedding*. It consists of taking each focal element $\mathsf{b} \in 2^{\Omega_Y}$ of $m_{Y|\mathsf{a}}$, and converting it to the corresponding focal element

$$(\mathsf{a} \times \mathsf{b}) \cup ((\Omega_X \setminus \mathsf{a}) \times \Omega_Y) \in 2^{\Omega_{X,Y}} \tag{7}$$

of $m_{\mathsf{a},Y}$ with the same mass. It is easy to confirm that this method of embedding satisfies both conditions mentioned above.

When does a belief function qualify as a conditional? For example, suppose we have a BPA m for $\{Y\} \cup \mathcal{X}$ where $\{Y\} \cap \mathcal{X} = \emptyset$. Under what conditions does m constitute a conditional for Y given $\mathcal{X}$? Analogous to conditional probability tables in Bayesian networks, the answer is straightforward. Any BPA m for $\{Y\} \cup \mathcal{X}$ such that $m^{\downarrow \mathcal{X}}$ is the vacuous BPA for $\mathcal{X}$ constitutes a conditional for Y given $\mathcal{X}$. Sometimes, we will let $m_{Y|\mathcal{X}}$ denote such conditionals.

3 Decomposable Entropy of D-S Belief Functions

This section reviews the definitions of decomposable entropy and conditional decomposable entropy of belief functions in the D-S theory (Jiroušek and Shenoy, 2018a) and describes its properties (Jiroušek and Shenoy, 2020). We also describe a new property of decomposable entropy motivated by the need to compute the decomposable entropy of an undirected graphical belief function model.

3.1 Decomposable Entropy

Definition 1 (Entropy of a CF Q) *Suppose Q is a CF for $\mathcal{X}$ with state-space $\Omega_{\mathcal{X}}$. Then, the decomposable entropy of Q, denoted by $H(Q)$, is defined as*

$$H(Q) = \sum_{a \in 2^{\Omega_{\mathcal{X}}}} (-1)^{|\mathsf{a}|} Q(a) \log(Q(a)). \tag{8}$$

The definition of entropy of Q in Eq. (8) is well-defined as it follows from the definition of a CF in Eq. (3) that for all $\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}$, $Q(\mathsf{a}) \geq 0$. If $Q(\mathsf{a}) = 0$, we will follow the convention that $Q(\mathsf{a}) \log(Q(\mathsf{a})) = 0$ as $\lim_{\theta \to 0^+} \theta \log(\theta) = 0$. Thus, in computing the entropy $H(Q)$ as defined in Def. 1, it is sufficient that the summation in the right-hand side of Eq. (8) is restricted to $\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}$ such that $Q(\mathsf{a}) > 0$.

3.2 Conditional Decomposable Entropy

Definition 2 (Conditional entropy of $Q_{Y|X}$) *Suppose Q_X is a CF for X, and suppose $Q_{Y|X}$ is a conditional CF for (X, Y). Then, the conditional decomposable entropy of $Q_{Y|X}$, denoted by $H(Q_{Y|X})$, is defined as follows:*

$$H(Q_{Y|X}) = \sum_{\mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|}\, Q_X(\mathsf{a}^{\downarrow X})\, Q_{Y|X}(\mathsf{a})\, \log(Q_{Y|X}(\mathsf{a})). \tag{9}$$

Notice that as $Q_X(\mathsf{a}^{\downarrow X})\, Q_{Y|X}(\mathsf{a}) = Q_{X,Y}(\mathsf{a})$ for all $\mathsf{a} \in 2^{\Omega_{X,Y}}$, we can rewrite Eq. (9) as follows:

$$H(Q_{Y|X}) = \sum_{\mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|}\, Q_{X,Y}(\mathsf{a})\, \log(Q_{Y|X}(\mathsf{a})) \tag{10}$$

3.3 Properties of Decomposable Entropy

A list of relevant properties of the decomposable entropy is as follows. For formal proofs, see (Jiroušek and Shenoy, 2020).

Property 1 (Compound distributions) *Suppose Q_X is a CF for X, and suppose $Q_{Y|X}$ is a conditional CF for (X, Y). Let $Q_{X,Y} = Q_X \oplus Q_{Y|X}$. Then,*

$$H(Q_{X,Y}) = H(Q_X) + H(Q_{Y|X}). \tag{11}$$

Property 2 (Quasi-consonant BPAs have 0 decomposable entropy) *Suppose m is a quasi-consonant BPA. Then H(m) = 0. As vacuous, deterministic, and consonant BPAs are also quasi-consonant, their decomposable entropies are also 0.*

Suppose P_X is a probability mass function (PMF) for X such that $P_X(x) > 0$ for all $x \in \Omega_X$, and $P_{Y|X}$ is a conditional probability table (CPT) for Y given X, i.e., $P_{Y|X}(x, y) = P_{Y|x}(y)$, where $P_{Y|x}$ is the conditional PMF for Y given $X = x$ for all $(x, y) \in \Omega_{X,Y}$. Let $P_{X,Y} = P_X \otimes P_{Y|X}$ ($\otimes$ denotes probabilistic combination, which is pointwise multiplication followed by normalization). Let m_X denote the Bayesian BPA corresponding to P_X, let $m_{Y|x}$ denote the Bayesian conditional BPA for Y corresponding to the conditional PMF $P_{Y|x}$ for Y given $X = x$. Let $m_{x,Y}$ denote the conditional BPA for (X, Y) obtained by conditional embedding of $m_{Y|x}$. Let $m_{Y|X}$ denote $\bigoplus_{x \in \Omega_X} m_{x,Y}$. Let $m_{X,Y}$ denote $m_X \oplus m_{Y|X}$. Notice that $m_{x,Y}$ and $m_{Y|X}$ are not Bayesian BPAs.

Property 3 (Strong probability consistency) *Consider the situation described in the preceding paragraph. Let $H_s(P_{X,Y})$ and $H_s(P_X)$ denote Shannon's entropy of PMFs $P_{X,Y}$ and P_X, respectively, and let $H_s(P_{Y|X})$ denote Shannon's conditional entropy of the CPT $P_{Y|X}$. Then, $m_{X,Y}$ is a Bayesian BPA for (X, Y) corresponding to PMF $P_{X,Y}$ such that:*

$$H(m_{X,Y}) = H_s(P_{X,Y}), \tag{12}$$
$$H(m_X) = H_s(P_X), \tag{13}$$
$$H(m_{Y|X}) = H_s(P_{Y|X}). \tag{14}$$

The following theorem generalizes Property 1. It is a new property not discussed in (Jiroušek and Shenoy, 2020). It is motivated by the need to compute the entropy of an undirected belief function graphical model.

Theorem 1 (Entropy of non-informative belief functions) *Suppose Q_1 and Q_2 are distinct CFs for $\mathcal{X}_1$, and $\mathcal{X}_2$, respectively, such that they are non-informative for each other. Then,*

$$H(Q_1 \oplus Q_2) = H(Q_1) + H(Q_2) \tag{15}$$

A proof of this property can be found in a longer version of this paper (Jiroušek et al., 2022).

4 An Algorithm

This section describes an algorithm for computing the decomposable entropy of a directed graphical belief function.

Suppose we have a directed acyclic graph G consisting of a set of variables $\{X_1, \ldots, X_n\}$ as nodes, and a set of directed edges. Let $Pa_G(X_k)$ denote the parents of X_k in graph G. Associated with each node X_k is a conditional BPA m_k for $X_k \cup Pa_G(X_k)$ that is a conditional for X_k given $Pa_G(X_k)$. If $Pa_G(X_k) = \emptyset$, then the conditional for X_k is the prior belief function for X_k. If $Pa_G(X_k) \neq \emptyset$, then we will assume that m_k is a conditional BPA for $X_k \cup Pa_G(X_k)$, i.e., $m_k^{\downarrow X_k}$ is a vacuous BPA for X_k.

Notice that if we have evidence for a variable that is different from priors or conditionals in a directed graphical belief function model, we need to disregard such evidence. For example, suppose we have a directed acyclic graph $X \to Y$ with a BPA m_1 for X, a conditional BPA m_2 for $\{X, Y\}$ that constitutes a conditional for $Y|X$ so that $m_2^{\downarrow X}$ is the vacuous BPA for X, and a BPA m_3 for Y that represents some evidence for Y. It follows from the compound distributions property that $H(m_1 \oplus m_2) = H(m_1) + H(m_2)$. But, in general, $H(m_1 \oplus m_2 \oplus m_3) \neq H(m_1) + H(m_2) + H(m_3)$. For this reason, we need to disregard evidence in computing the decomposable entropy of a directed graphical belief function model.

Algorithm First, we start with a sequence $(X_1, \ldots X_n)$ such that if there is a directed arc $X_i \to X_j$ in G, then X_i precedes X_j in the sequence. As G is acyclic, such a sequence always exists, but it may not be unique.
Do $k = 1, \ldots, n$:

- If $Pa_G(X_k) = \emptyset$, then $H(m_k)$ is computed using Definition 1.
- If $Pa_G(X_k) \neq \emptyset$, then first we find the marginal $(\bigoplus_{i=1}^{k-1} m_i)^{\downarrow Pa_G(X_k)}$ using local computation (Shenoy and Shafer, 1990). Next, we find the conditional decomposable entropy of m_k, $H(m_k)$, using Definition 2.

End Do;

The decomposable entropy of the joint belief function $H(\bigoplus_{k=1}^{n} m_k) = \sum_{k=1}^{n} H(m_k)$. This follows from the compound distributions property of decomposable entropy.

5 Three Examples

This section computes the decomposable entropy of three graphical belief function models.

Captain's Problem The captain's problem is from Almond (1995). A ship's captain is concerned about how many days his ship may be delayed before arrival at a destination. The arrival delay is the sum of departure delay and sailing delay. Departure delay may be a result of maintenance (at most one day), loading delay (at most one day), or a forecast of bad weather (at most one day). Sailing delays may result from bad weather (at most one day) and whether repairs are needed at sea (at most one day). If maintenance is done before sailing, chances of repairs at sea are less likely. The weather forecast says a slight chance of bad weather (0.2) and a good chance of good weather (0.6). The forecast is 80% reliable. The captain knows the loading delay and whether maintenance is done before departure. Fig. 1 shows the directed acyclic graph associated with this problem. Table 1 shows the variables, their state spaces, and the associated conditionals. What is the decomposable entropy of this belief function model?

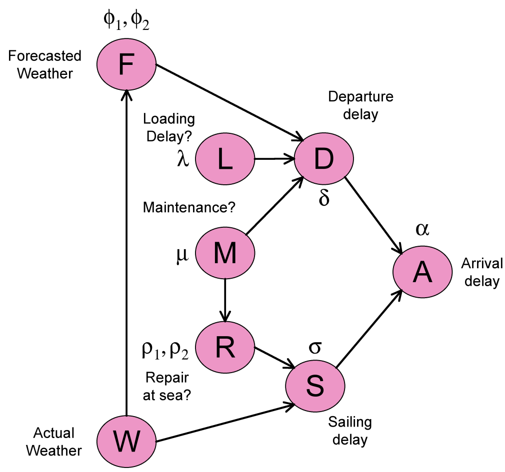


Figure 1: The directed acyclic graph for the captains's problem. The Greek alphabets adjacent to a variable denote the prior or conditional or evidence associated with the variable.

As ϕ_2 is an evidence for F, we ignore this belief function. First, notice that ϕ_1 and σ are consonant, and μ, δ, and α are deterministic. So the decomposable entropies of these BPAs are zeroes. The decomposable entropies of the remaining BPAs are as follows.

Table 1: The variables, their state spaces, and associated conditionals in the captain's problem.

Variable	Name	State Space, Ω	Associated Conditional
W	Actual weather	$\{g_w, b_w\}$	vacuous for W
F	Forecasted weather	$\{g_f, b_f\}$	ϕ_1 for $F\|W$ (consonant)
L	Loading delay?	$\{t_l, f_l\}$	λ for L
M	Maintenance done?	$\{t_m, f_m\}$	μ for M (deterministic)
R	Repair at sea needed?	$\{t_r, f_r\}$	$\rho_1 \oplus \rho_2$ for $R\|M$
D	Departure delay (in days)	$\{0, 1, 2, 3\}$	δ for $D\|\{F, L, M\}$ (deterministic)
S	Sailing delay (in days)	$\{0, 1, 2, 3\}$	σ for $S\|\{W, R\}$ (consonant)
A	Arrival delay (in days)	$\{0, 1, 2, 3, 4, 5, 6\}$	α for $A\|\{D, S\}$ (deterministic)

$H(\lambda) \approx 0.3958$, $H(\rho_1 \oplus \rho_2) \approx 0.0729$, Thus, the decomposable entropy of the captain's problem (ignoring the evidence ϕ_2) is $0.3958 + 0.0729 = 0.4687$.

Communication Network This example is from Haenni and Lehmann (2002). Fig. 2 shows an undirected graph associated with this example. We have a grid of $44 = 8 + 9 + 10 + 9 + 8$ communication nodes arranged in 19 columns and 5 rows. There are 68 links, and each link has 90% reliability. Nodes A and B are connected to the grid with links having 80% reliability. What is the decomposable entropy of this graphical model?

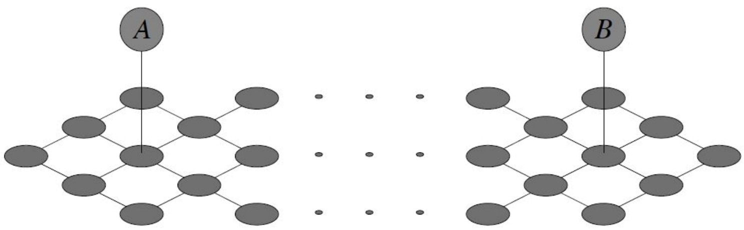


Figure 2: The undirected graph for the communication network example.

Consider the variables in the grid with 19 columns and 5 rows. Let X_{13} denote the variable in column 1, row 3 and let X_{22} denote the variable in column 2 and row 2. Let $\Omega_{13} = \{t_{13}, f_{13}\}$, and and let $\Omega_{22} = \{t_{22}, f_{22}\}$. The BPA m_{13-22} associated with the edge between X_{13} and X_{22} is as follows:

$$m_{13-22}(\{(t_{13}, t_{22}), (f_{13}, f_{22})\}) = 0.9, m_{13-22}(\Omega_{13} \times \Omega_{22}) = 0.1.$$

The BPAs associated with the remaining 67 links are similar. The edges between A and X_{33} and between B and X_{38} are also similar, except that the reliability is 0.8 instead of 0.9. As these BPAs are consonant, the decomposable entropy of all 70 BPAs are zeroes. Also, notice that the BPAs m_{13-22} and m_{12-24} associated with the corresponding edges satisfy the conditions in Theorem 1. As all the BPAs in this example have the same structure, the set of all BPAs is non-informative. Thus, the decomposable entropy of the communication network model is 0.

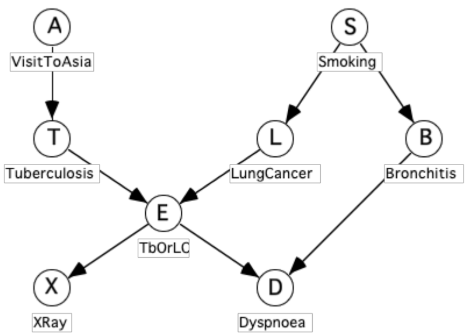
Chest Clinic This example is from Lauritzen and Spiegelhalter (1988). Fig. 3 shows a Bayesian network that is represented as a directed graphical belief function model. There are eight binary variables, and not all probabilities in the joint probability distribution are positive. Fig. 3 also shows the conditional probability tables (CPTs). These are represented as BPAs using conditional embedding, and most of these BPAs are not Bayesian. The decomposable entropies of the conditionals are as follows (computed using the algorithm in Section 4):

$$H(P(A)) \approx 0.0808, H(P(T|A)) \approx 0.0828, H(P(S)) = 1, H(P(L|S)) \approx 0.2749,$$

$$H(P(B|S)) \approx 0.9261, H(P(E|L,T)) = 0, H(P(X|E)) \approx 0.2770, H(P(B|DE)) \approx 0.6471.$$

Thus, the decomposable entropy of the directed graphical belief function model is approximately 3.2887, which is the same as Shannon's entropy of the corresponding Bayesian network.

Figure 3: The directed acyclic graph and the CPTs for the chest clinic example.



P(A):	$p(a) = .01$	P(E|L,T):	$p(e \mid l, t) = 1$ $p(e \mid l, \sim t) = 1$ $p(e \mid \sim l, t) = 1$ $p(e \mid \sim l, \sim t) = 0$
P(T|A):	$p(t \mid a) = .05$ $p(t \mid \sim a) = .01$		
P(S):	$p(s) = .50$	P(X|E):	$p(x \mid e) = .98$ $p(x \mid \sim e) = .05$
P(L|S):	$p(l \mid s) = .10$ $p(l \mid \sim s) = .01$	P(D|E,B):	$p(d \mid e, b) = .90$ $p(d \mid e, \sim b) = .70$ $p(d \mid \sim e, b) = .80$ $p(d \mid \sim e, \sim b) = .10$
P(B|S):	$p(b \mid s) = .60$ $p(b \mid \sim s) = .30$		

6 Notes on Implementation

We performed all experiments in *R*. We have created an *R* package to work with belief functions, which we plan to complete and publish for use by other users. The package is based on relational databases as implemented in the *R* package *data.table* (Dowle and Srinivasan, 2021). Each belief function is an object with three different tables. The first table, called the *coding table*, consists of random variables and their states. The columns correspond to the random variables in the domain $\mathcal{X}$ of the belief function, the rows to the elements of their joint state space $\times_{X \in \mathcal{X}} \Omega_X$. Each row is labeled with a unique identifier. The second table, called *focal element table*, stores each focal element as a set of states using identifiers from the *coding table*. The third table, called *mass table*, assigns a probability mass to each focal element.

Regarding computing the marginal of the joint in the algorithm described in Section 4, an implementation using local computation is available in the *Belief Function Machine* environment in Matlab (Giang and Shenoy, 2003). We implemented this algorithm in *R*.

7 Summary & Conclusions

The primary goal is to describe an algorithm for computing the decomposable entropy of directed graphical belief function models. The decomposable entropy has a property that if we construct a joint BPA for two variables (X, Y) by Dempster's combination of a BPA for m_X for X and a conditional BPA $m_{Y|X}$ for Y given X, then the decomposable entropy of $m_{X,Y} = m_X \oplus m_{Y|X}$ is equal to the decomposable entropy of m_X plus the decomposable conditional entropy of $m_{Y|X}$.

The decomposable entropy is defined using commonality functions. If a graphical model has a clique whose state space is large, then computing the decomposable entropy of the clique may be intractable. For example, in the captain's problem, the conditional for arrival delay has three variables with a joint space of $4 \times 4 \times 7 = 112$ states. Fortunately, this conditional is deterministic, and the decomposable entropy of deterministic BPAs is 0. If this conditional wasn't deterministic or consonant or quasi-consonant, and the joint commonality function for these three variables had non-zero values for each of the 2^{112} subsets, then the computation of the exact decomposable entropy of the conditional would be intractable. In such cases, we may have to resort to some approximate methods. This is yet to be done.

Acknowledgement

This study was financially supported by the Czech Science Foundation under Grant No. 19-06569S (first author), grant GA ČR 19-04579S (second author), and by the Ronald G. Harper Professorship at the University of Kansas awarded to the third author.

References

R. G. Almond. *Graphical Belief Modeling.* Chapman & Hall, London, UK, 1995.

A. P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B*, 30(2):205–247, 1968.

M. Dowle and A. Srinivasan. *data.table: Extension of 'data.frame'*, 2021. URL `https://CRAN.R-project.org/package=data.table`. R package version 1.14.2.

P. Giang and S. Shenoy. *The belief function machine: An environment for reasoning with belief functions in Matlab.* Lawrence, KS 66045, 2003. URL `https://pshenoy.ku.edu/Papers/BFM072503.zip`.

R. Haenni and N. Lehmann. Resource bounded and anytime approximation of belief function computation. *International Journal of Approximate Reasoning*, 31(1–2):103–154, 2002.

R. Jiroušek and P. P. Shenoy. A decomposable entropy for belief functions in the Dempster-Shafer theory. In S. Destercke, T. Denoeux, F. Cuzzolin, and A. Martin, editors, *Belief Functions: Theory and Applications*, volume 11069 of *Lecture Notes in Artificial Intelligence*, pages 146–154. Springer Nature, Switzerland, 2018a.

R. Jiroušek and P. P. Shenoy. A new definition of entropy of belief functions in the Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 92(1):49–65, 2018b.

R. Jiroušek and P. P. Shenoy. On properties of a new decomposable entropy of Dempster-Shafer belief functions. *International Journal of Approximate Reasoning*, 119(4):260–279, 2020.

R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. Computing the decomposable entropy of graphical belief function models. Working Paper 340, University of Kansas School of Business, Lawrence, KS 66045, USA, 2022. URL `https://pshenoy.ku.edu/Papers/WP340.pdf`.

S. L. Lauritzen and D. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, series B*, 50(2):157–224, 1988.

G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

G. Shafer. Belief functions and parametric models. *Journal of the Royal Statistical Society, Series B*, 44(3):322–352, 1982.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

P. P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. In R. D. Shachter, T. Levitt, J. F. Lemmer, and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*, Machine Intelligence and Pattern Recognition Series, vol. 9, pages 169–198, Amsterdam, 1990. North-Holland.

P. Smets. *Un modele mathematico-statistique simulant le processus du diagnostic medical.* PhD thesis, Free University of Brussels, 1978.

H. Xu and P. Smets. Reasoning in evidential networks with conditional belief functions. *International Journal of Approximate Reasoning*, 14(2–3):155–185, 1996.

Two Composition Operators for Belief Functions Revisited

Radim Jiroušek[1,2], Václav Kratochvíl[1,2], and Prakash P. Shenoy[3]

[1]Faculty of Management, University of Economics, Jindřichův Hradec, Czechia
[2]Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czechia
[1,2] {*radim, velorex*}*@utia.cas.cz*
[3]School of Business, University of Kansas, Lawrence, KS, USA
[3] *pshenoy@ku.edu*

May 8, 2022

Abstract

In probability theory, compositional models are as powerful as Bayesian networks. However, the relation between belief-function graphical models and the corresponding compositional models is much more complicated due to several reasons. One of them is that there are two composition operators for belief functions. This paper deals with their main properties and presents sufficient conditions under which they yield the same results.

1 Introduction

Two different composition operators for belief functions are defined in the literature (Jiroušek et al., 2007; Jiroušek and Shenoy, 2014). Surprisingly, for directed graphical belief function models, e.g., Almond's 'Captain's problem' (Almond, 1995), the corresponding compositional models are the same (regardless of the operator used). This unexpected finding is surprising since the two operators are designed based on different ideas and for different purposes. Historically, the first operator (Jiroušek et al., 2007), called the *f-composition* operator here, is designed to represent multivariate basic probabilistic assignments (BPA) using the lowest number of parameters. The second operator (Jiroušek et al., 2007) is consistent with the Dempster-Shafer (D-S) theory of evidence, and therefore, we call it the *d-composition* operator. The d-composition operator introduces conditional independence relations among the variables, similar to the probabilistic composition operator. Thus, the class of d-compositional models is equivalent to the class of directed probabilistic graphical models. In general, this is not true for f-compositional models.

The idea behind the f-composition operator is to decrease the number of parameters necessary for representing multidimensional belief functions. For example, instead of representing one three-dimensional BPA, one represents only two two-dimensional BPAs. In the case of binary variables it means that one can use only $2 \times 2^{(2^2)} = 32$ instead of $2^{(2^3)} = 128$ parameters. Naturally, there is no free lunch, and one has to pay for it by restricting the class of such BPAs. One has to give up the possibility of using belief functions whose BPAs are not factorizable.

In probability theory, there is a *factorization lemma*[1] that says if a joint probability distribution P for variables X, Y, and Z can be expressed in the form of a product of two factors $\phi_1(X, Y)$ and $\phi_2(Y, Z)$, then X and Z are conditionally independent given Y, written as $X \perp\!\!\!\perp Z | Y$. Therefore, for a probabilistic compositional model, one can identify the induced conditional independence relations based on the factors in the model. There is a similar result for the D-S theory (Shenoy, 1994) and thus, for the d-composition operator. However, it is not clear what belief function theory corresponds to the f-composition operator, and therefore the problem of identification of the conditional independence relations for f-composition models is not obvious. However, there are other problems associated with the d-composition operator. The result of d-composition is sometimes undefined (see Definition 2 in Section 3).

Thus, it is not clear which composition operator is better. The user should choose the one which suits better the purpose of the application. Both of them satisfy the properties expected from composition operators (described in Section 3). Both composition operators have corresponding inverse decomposition operators. One of them corresponds to the notion of conditional independence, the other to a specific way of factorization. We will not study decompositions explicitly in this paper.

An outline of the remainder of the paper is as follows. Section 2 introduce the necessary concepts and notation from belief function theory. Section 3 contains definitions of the two composition operators. Section 4 has the main result of this paper. Section 5 illustrates the main result using Almond's captain's problem (Almond, 1995). Section 6 has a summary and some concluding remarks.

2 Belief Functions

Let $\mathcal{W}$ denote a set of variables each with finite number of states. For $X \in \mathcal{W}$, let Ω_X denote the set of states of variable X. A *basic probability assignment* (BPA) for variables $\mathcal{U} \subseteq \mathcal{W}$ (or equivalently, a BPA on the Cartesian product $\Omega_{\mathcal{U}} = \times_{X \in \mathcal{U}} \Omega_X$) is a mapping $m_{\mathcal{U}} : 2^{\Omega_{\mathcal{U}}} \to [0, 1]$, such that $\sum_{\mathbf{a} \subseteq \Omega_{\mathcal{U}}} m_{\mathcal{U}}(\mathbf{a}) = 1$ and $m_{\mathcal{U}}(\emptyset) = 0$.

Consider a BPA $m_{\mathcal{U}}$. If the set of the corresponding variables is clear from the context, we omit the subscript $\mathcal{U}$. We say $\mathbf{a}$ is a *focal element* of m if $m(\mathbf{a}) > 0$. If m has only one

[1] It is not the same as the factorization lemma from the theory of categories.

focal element, we say m is *deterministic*. If this focal element is $\Omega_{\mathcal{U}}$, i.e., $m(\Omega_{\mathcal{U}}) = 1$, we say that m is *vacuous*.

Given a BPA m, the same information can be expressed by the corresponding *commonality* function (which is also defined on the power set 2^{Ω}):

$$Q_m(\mathbf{a}) = \sum_{\mathbf{b} \subseteq \Omega:\, \mathbf{b} \supseteq \mathbf{a}} m(\mathbf{b}). \tag{1}$$

Whenever a commonality function is given, it is possible to reconstruct the corresponding BPA m:

$$m(\mathbf{a}) = \sum_{\mathbf{b} \subseteq \Omega: \mathbf{b} \supseteq \mathbf{a}} (-1)^{|\mathbf{b} \setminus \mathbf{a}|} Q_m(\mathbf{b}). \tag{2}$$

For BPA $m_{\mathcal{V}}$, we often consider its *marginal* $m_{\mathcal{V}}^{\downarrow \mathcal{U}}$ for $\mathcal{U} \subseteq \mathcal{V}$. A similar notation is used also for *projections* of states. If $a \in \Omega_{\mathcal{V}}$, $a^{\downarrow \mathcal{U}}$ denotes the element of $\Omega_{\mathcal{U}}$, which is obtained from a by omitting the values of variables in $\mathcal{V} \setminus \mathcal{U}$. For $\mathbf{a} \subseteq \Omega_{\mathcal{V}}$,

$$\mathbf{a}^{\downarrow \mathcal{U}} = \{a^{\downarrow \mathcal{U}} : a \in \mathbf{a}\}.$$

Using this notation, the marginal $m_{\mathcal{V}}^{\downarrow \mathcal{U}}$ of BPA $m_{\mathcal{V}}$ for $\mathcal{U} \subseteq \mathcal{V}$ is defined as follows:

$$m_{\mathcal{V}}^{\downarrow \mathcal{U}}(\mathbf{b}) = \sum_{\mathbf{a} \subseteq \Omega_{\mathcal{V}}:\, \mathbf{a}^{\downarrow \mathcal{U}} = \mathbf{b}} m_{\mathcal{V}}(\mathbf{a}).$$

for all $\mathbf{b} \subseteq \Omega_{\mathcal{U}}$.

The projection of sets enables us to define a *join* of two sets. Consider two arbitrary sets $\mathcal{U}$ and $\mathcal{V}$ of variables (they may be disjoint or overlapping, or one may be a subset of the other). Consider two sets $\mathbf{a} \subseteq \Omega_{\mathcal{U}}$ and $\mathbf{b} \subseteq \Omega_{\mathcal{V}}$. Their join is defined as

$$\mathbf{a} \bowtie \mathbf{b} = \{c \in \Omega_{\mathcal{U} \cup \mathcal{V}} : c^{\downarrow \mathcal{U}} \in \mathbf{a} \;\;\&\;\; c^{\downarrow \mathcal{V}} \in \mathbf{b}\}.$$

Notice that if $\mathcal{U}$ and $\mathcal{V}$ are disjoint, then $\mathbf{a} \bowtie \mathbf{b} = \mathbf{a} \times \mathbf{b}$. If $\mathcal{U} = \mathcal{V}$, then $\mathbf{a} \bowtie \mathbf{b} = \mathbf{a} \cap \mathbf{b}$. In general, for $\mathbf{c} \subseteq \Omega_{\mathcal{U} \cup \mathcal{V}}$, $\mathbf{c}$ is a subset of $\mathbf{c}^{\downarrow \mathcal{U}} \bowtie \mathbf{c}^{\downarrow \mathcal{V}}$, which may be a proper subset. If $\mathbf{c}^{\downarrow \mathcal{U} \cap \mathcal{V}}$ is a singleton subset, then $\mathbf{c} = \mathbf{c}^{\downarrow \mathcal{U}} \bowtie \mathbf{c}^{\downarrow \mathcal{V}}$.

To construct multidimensional models from low-dimensional building blocks, we need some operators connecting two low-dimensional BPAs into one BPA. One possibility is the classical Dempster's combination rule, which is used to combine distinct belief functions. Consider two BPAs $m_{\mathcal{U}}$ and $m_{\mathcal{V}}$ for arbitrary sets of variables $\mathcal{U}$ and $\mathcal{V}$. Dempster's combination, denoted by $\oplus$, is defined as follows (Shafer, 1976):

$$(m_{\mathcal{U}} \oplus m_{\mathcal{V}})(\mathbf{c}) = \frac{1}{1-K} \sum_{\mathbf{a} \subseteq \Omega_{\mathcal{U}}, \mathbf{b} \subseteq \Omega_{\mathcal{V}} : \mathbf{a} \bowtie \mathbf{b} = \mathbf{c}} m_{\mathcal{U}}(\mathbf{a}) \cdot m_{\mathcal{V}}(\mathbf{b}), \tag{3}$$

for each $\mathbf{c} \subseteq \Omega_{\mathcal{U} \cup \mathcal{V}}$, where

$$K = \sum_{\mathbf{a} \subseteq \Omega_{\mathcal{U}}, \mathbf{b} \subseteq \Omega_{\mathcal{V}} : \mathbf{a} \bowtie \mathbf{b} = \emptyset} m_{\mathcal{U}}(\mathbf{a}) \cdot m_{\mathcal{V}}(\mathbf{b}). \tag{4}$$

K can be interpreted as the amount of conflict between $m_{\mathcal{U}}$ and $m_{\mathcal{V}}$. If $K = 1$, we say $m_{\mathcal{U}}$ and $m_{\mathcal{V}}$ are in *total conflict* and their Dempster's combination is undefined.

3 Composition Operator

The following definition answers the question: What do we mean by a belief function composition operator?

Definition 1 *By an* composition operator $\triangleright$ *we mean any binary operator satisfying the following four axioms. Consider arbitrary three BPAs $m_{\mathcal{T}}$, $m_{\mathcal{U}}$, and $m_{\mathcal{V}}$.*

A1 (Domain): *$m_{\mathcal{T}} \triangleright m_{\mathcal{U}}$ is a BPA for $\mathcal{T} \cup \mathcal{U}$.*

A2 (Composition preserves first marginal): $(m_{\mathcal{T}} \triangleright m_{\mathcal{U}})^{\downarrow \mathcal{T}} = m_{\mathcal{T}}$.

A3 (Commutativity under consistency): *If $m_{\mathcal{T}}$ and $m_{\mathcal{U}}$ are consistent, i.e., $m_{\mathcal{T}}^{\downarrow \mathcal{T} \cap \mathcal{U}} = m_{\mathcal{U}}^{\downarrow \mathcal{T} \cap \mathcal{U}}$, then $m_{\mathcal{T}} \triangleright m_{\mathcal{U}} = m_{\mathcal{U}} \triangleright m_{\mathcal{T}}$.*

A4 (Restricted associativity): *If $\mathcal{T} \supset (\mathcal{U} \cap \mathcal{V})$, or, $\mathcal{U} \supset (\mathcal{T} \cap \mathcal{V})$, then $(m_{\mathcal{T}} \triangleright m_{\mathcal{U}}) \triangleright m_{\mathcal{V}} = m_{\mathcal{T}} \triangleright (m_{\mathcal{U}} \triangleright m_{\mathcal{V}})$.*

Notice that axioms *A1*, *A3*, *A4* guarantee that the composition operator uniquely reconstructs BPA $m_{\mathcal{T} \cup \mathcal{V}}$ from its marginals, if there exists a *lossless* decomposition of $m_{\mathcal{T} \cup \mathcal{V}}$ into $m_{\mathcal{T}}$ and $m_{\mathcal{V}}$. Surprisingly, it is axiom *A4*, which guarantees that no necessary information from $m_{\mathcal{V}}$ is lost. Axiom *A2* solves the problem arising when non-consistent basic assignments are composed. Generally, there are two ways of coping with this problem. Either find a compromise (a mixture of inconsistent pieces of knowledge) or give preference to one of the sources. The solution expressed by axiom *A2* is superior to the other two from a computational point of view.

The following assertion (for proofs see (Jiroušek and Shenoy, 2014)) summarizes the main properties of composition operators. Based on these, efficient computational procedures were designed.

Proposition 1 *For arbitrary BPAs $m_{\mathcal{T}}, m_{\mathcal{U}}, m_{\mathcal{V}}$ the following statements hold.*

1. (Reduction:) *If $\mathcal{U} \subseteq \mathcal{T}$, then $m_{\mathcal{T}} \triangleright m_{\mathcal{U}} = m_{\mathcal{T}}$.*
2. (Stepwise composition): *If $(\mathcal{T} \cap \mathcal{U}) \subseteq \mathcal{V} \subseteq \mathcal{U}$, then $(m_{\mathcal{T}} \triangleright m_{\mathcal{U}}^{\downarrow \mathcal{V}}) \triangleright m_{\mathcal{U}} = m_{\mathcal{T}} \triangleright m_{\mathcal{U}}$.*
3. (Exchangeability): *If $\mathcal{U} \supset (\mathcal{T} \cap \mathcal{V})$, then $(m_{\mathcal{T}} \triangleright m_{\mathcal{U}}) \triangleright m_{\mathcal{V}} = (m_{\mathcal{T}} \triangleright m_{\mathcal{V}}) \triangleright m_{\mathcal{U}}$.*
4. (Simple marginalization): *If $(\mathcal{T} \cap \mathcal{U}) \subseteq \mathcal{V} \subseteq (\mathcal{T} \cup \mathcal{U})$, then $(m_{\mathcal{T}} \triangleright m_{\mathcal{U}})^{\downarrow \mathcal{V}} = m_{\mathcal{T}}^{\downarrow \mathcal{T} \cap \mathcal{V}} \triangleright m_{\mathcal{U}}^{\downarrow \mathcal{U} \cap \mathcal{V}}$.*

Before defining a composition operator for the D-S theory, notice that Dempster's combination rule is *not* a composition operator. Though it satisfies the first axiom (Domain), it does not satisfy the remaining three axioms. Whereas Dempster's rule is commutative and associative, a composition operator only satisfies these properties in special situations. On the other hand, Dempster's rule does not preserve the first marginal. Dempster's rule is designed to combine distinct pieces of evidence, whereas composition is designed to combine marginals that may not be independent. Nevertheless, as shown below, Dempster's rule can be used to define a composition operator.

3.1 d-composition

In this paper we follow the idea introduced in (Jiroušek and Shenoy, 2018). It defines d-composition of two BPAs $m_{\mathcal{U}}$, $m_{\mathcal{V}}$ (for any $\mathcal{U}, \mathcal{V}$) as follows.

$$(m_{\mathcal{U}} \triangleright_d m_{\mathcal{V}}) = m_{\mathcal{U}} \oplus m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}},$$

where $\ominus$ denotes the inverse to Dempster's combination rule. $\ominus$ is defined using the corresponding commonality functions. Since it is known that the Dempster's rule can be stated as the product of the corresponding commonality functions (Shafer (1976)), i.e.,

$$Q_{m_1 \oplus m_2} = \frac{1}{1-K}\, Q_{m_1} \cdot Q_{m_2},$$

where K is the normalization factor from Eq. (3) defined by Eq. (4). Thus, $m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}$ was computed as a BPA corresponding to the following commonality function

$$Q_{m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}} = \frac{Q_{m_{\mathcal{V}}}}{Q_{m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}}}.$$

However, as shown in (Jiroušek and Shenoy, 2014), the composition operator $\triangleright_d$ sometimes yields BPAs with negative values (such BPAs are often called pseudo-BPAs).

Example 1 Consider $\Omega_X = \{x, \bar{x}\}$, $\Omega_Y = \{y, \bar{y}\}$, which means that $|2^{\Omega_X}| = 4$, and $|2^{\Omega_{XY}}| = 16$. In this example, consider a BPA m_{XY} for (X, Y) with only two focal elements – see Table 1. In tables, we depict only focal elements, i.e., if $\mathbf{a} \subseteq \Omega$ is not included in the table, then its value is 0.

Table 1: A Simple Example m_{XY}

$\mathbf{a}$	$m_{XY}(\mathbf{a})$
$\{(x, y)\}$	0.9
$\{(x, y), (x, \bar{y}), (\bar{x}, \bar{y})\}$	0.1

Its marginal $m_X = m_{XY}^{\downarrow X}$ has also two focal elements, namely $m_X(\{x\}) = 0.9$ and $m_X(\Omega_X) = 0.1$. Therefore, the corresponding commonality function is as follows: $Q_{m_X}(\{x\}) = 1$, $Q_{m_X}(\{\bar{x}\}) = Q_{m_X}(\Omega_X) = 0.1$. The computation of the corresponding $Q_{m_{XY} \ominus m_X}$ and $m_{XY} \ominus m_X$ can be seen in Table 2. □

To avoid situations when the result of a composition is not a BPA, in this paper, we accept the possibility that the result of the operation of composition is undefined. Another advantage of this approach is that we also avoid the necessity of using commonality functions.

Definition 2 *Suppose $m_{\mathcal{U}}$ and $m_{\mathcal{V}}$ are BPAs. If $m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}$ is a BPA, then the d-composition is defined as follows:*

$$m_{\mathcal{U}} \triangleright_d m_{\mathcal{V}} = m_{\mathcal{U}} \oplus (m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}). \tag{5}$$

If $m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}$ is not a BPA, then $m_{\mathcal{U}} \triangleright_d m_{\mathcal{V}}$ is undefined.

Table 2: Computation of $(m_{XY} \ominus m_X)(\mathbf{a})$.

$\mathbf{a}$	$Q_{m_{XY}}(\mathbf{a})$	$Q_{m_X}(\mathbf{a}^{\downarrow X})$	$Q_{m_{XY} \ominus m_X}(\mathbf{a}) = \frac{Q_{m_{XY}}(\mathbf{a})}{Q_{m_X}(\mathbf{a}^{\downarrow X})}$	$(m_{XY} \ominus m_X)(\mathbf{a})$
$\{(x,y)\}$	1	1	1	0.9
$\{(x,\bar{y})\}$	0.1	1	0.1	
$\{(\bar{x},\bar{y})\}$	0.1	0.1	1	
$\{(x,y),(x,\bar{y})\}$	0.1	1	0.1	-0.9
$\{(x,y),(\bar{x},\bar{y})\}$	0.1	0.1	1	
$\{(x,\bar{y}),(\bar{x},\bar{y})\}$	0.1	0.1	1	
$\Omega_{X,Y} \setminus \{(\bar{x},y)\}$	0.1	0.1	1	1

Remark 1 A disadvantage of this definition follows from the fact that neither the axioms of Definition 1, nor Properties expressed in Proposition 1 generally hold exactly as they are expressed. Namely, one has to add that they hold under the assumption that the corresponding compositions are defined. As an example, consider the Stepwise composition (Property 2 from Proposition 1) with $\mathcal{T} = \emptyset$: If $\mathcal{U} \subseteq \mathcal{V}$, then $m_{\mathcal{V}}^{\mathcal{U}} \triangleright m_{\mathcal{V}} = m_{\mathcal{V}}$. Naturally, this equality can hold only when $m_{\mathcal{V}}^{\mathcal{U}} \triangleright m_{\mathcal{V}}$ is defined.

3.2 f-composition

The f-composition operator is defined as follows:

Definition 3 *Consider two BPAs $m_{\mathcal{U}}$ and $m_{\mathcal{V}}$. Their f-composition is a BPA $m_{\mathcal{U}} \triangleright_f m_{\mathcal{V}}$ defined for each nonempty $\mathbf{c} \subseteq \Omega_{\mathcal{U} \cup \mathcal{V}}$ by one of the following expressions:*

(i) *If $m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}(\mathbf{c}^{\downarrow \mathcal{U} \cap \mathcal{V}}) > 0$ and $\mathbf{c} = \mathbf{c}^{\downarrow \mathcal{U}} \bowtie \mathbf{c}^{\downarrow \mathcal{V}}$, then* $(m_{\mathcal{U}} \triangleright_f m_{\mathcal{V}})(\mathbf{c}) = \dfrac{m_{\mathcal{U}}(\mathbf{c}^{\downarrow \mathcal{U}}) \cdot m_{\mathcal{V}}(\mathbf{c}^{\downarrow \mathcal{V}})}{m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}(\mathbf{c}^{\downarrow \mathcal{U} \cap \mathcal{V}})}$;

(ii) *If $m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}(\mathbf{c}^{\downarrow \mathcal{U} \cap \mathcal{V}}) = 0$ and $\mathbf{c} = \mathbf{c}^{\downarrow \mathcal{U}} \times \Omega_{\mathcal{V} \setminus \mathcal{U}}$, then $(m_{\mathcal{U}} \triangleright_f m_{\mathcal{V}})(\mathbf{c}) = m_{\mathcal{U}}(\mathbf{c}^{\downarrow \mathcal{U}})$;*

(iii) *In all other cases, $(m_{\mathcal{U}} \triangleright_f m_{\mathcal{V}})(\mathbf{c}) = 0$.*

Remark 2 f-composition is always defined. Notice that if $m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}(\mathbf{c}^{\downarrow \mathcal{U} \cap \mathcal{V}}) = 0$ (i.e., the formula in case (**i**) is undefined), then the definition accepts a heuristic solution saying "I do not know".

4 Properties of Composition Operators

First, we prove the following simple assertion characterizing $m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}}$. A similar result is stated by Shenoy (1994) in the context of valuation-based systems.

Proposition 2 *Consider nonempty sets of variables $\mathcal{U} \subsetneq \mathcal{V}$ and BPA $m_{\mathcal{V}}$. If $(m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow \mathcal{U} \cap \mathcal{V}})$ is a BPA, then the following two properties hold:*

- $m_{\mathcal{V}} = m_{\mathcal{V}}^{\downarrow\mathcal{U}} \oplus (m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow\mathcal{U}\cap\mathcal{V}})$;
- $\left(m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow\mathcal{U}\cap\mathcal{V}}\right)^{\downarrow\mathcal{U}}$ *is vacuous.*

Proof. The first property is a direct implication of the associativity and commutativity of the Dempster's rule of combination, and the latter one follows immediately from the property

$$\mathcal{W} \supseteq \mathcal{T} \supseteq \mathcal{W} \cap \mathcal{V} \implies (m_{\mathcal{V}} \oplus m_{\mathcal{W}})^{\downarrow\mathcal{T}} = m_{\mathcal{V}} \oplus m_{\mathcal{W}}^{\downarrow\mathcal{T}}$$

called "local computation" (Shenoy and Shafer, 1990). □

These two properties are often expected to hold for the conditional BPA $m_{\mathcal{V}\setminus\mathcal{U}|\mathcal{U}}$. Recall that the conditional BPA was defined by Smets (1978) and Shafer (1982) using so-called conditional embedding. We do not need this notion in this paper, and so we do not present the definition. Nevertheless, it may be an interesting question for the future study to find out under what conditions $m_{\mathcal{V}\setminus\mathcal{U}|\mathcal{U}} = m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow\mathcal{U}\cap\mathcal{V}}$.

In Example 1 we presented a simple BPA m_{XY} for which $m_{XY} \ominus m_X$ was not a BPA. It means that there are BPAs that cannot be a second argument of a *d*-composition. From this, however, one cannot exclude the existence of another BPA for which the properties from Proposition 2 hold. Thus, let us turn back to the above-presented example and show that for m_{XY} from Table 1 such a BPA does not exist.

Example 1 (***Continued.***) Let us assume that there exists two-dimensional BPA $m_{Y|X}$ such that for m_{XY} from Table 1 $m_{XY} = m_{XY}^{\downarrow X} \oplus m_{Y|X}$. Then, under this assumption, for all $\mathbf{a} \subseteq \Omega_{XY}$

$$m_{XY}(\mathbf{a}) = (1-K)^{-1} \sum_{\mathbf{b}\subseteq\Omega_X \& \mathbf{c}\subseteq\Omega_{XY}:\mathbf{b}\bowtie\mathbf{c}=\mathbf{a}} m_{XY}^{\downarrow X}(\mathbf{b}) \cdot m_{Y|X}(\mathbf{c}). \tag{6}$$

Since $\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\} = \{x,\bar{x}\} \bowtie \{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}$, and for no other $\mathbf{b},\mathbf{c}$ their join $\mathbf{b} \bowtie \mathbf{c} = \{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}$, it is clear that $m_{Y|X}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}) = (1-K)$ because

$$\begin{aligned}
0.1 &= m_{XY}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}) \\
&= (1-K)^{-1}\, m_{XY}^{\downarrow X}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}^{\downarrow X}) \cdot m_{Y|X}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}) \\
&= (1-K)^{-1} \cdot m_{XY}^{\downarrow X}(\{x,\bar{x}\}) \cdot m_{Y|X}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}) \\
&= (1-K)^{-1} \cdot 0.1 \cdot m_{Y|X}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}).
\end{aligned}$$

Since $\{x\} \bowtie \{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\} = \{(x,y),(x,\bar{y})\}$, it immediately follows from (6) that

$$\begin{aligned}
m_{XY}(\{(x,y),(x,\bar{y})\}) &\geq (1-K)^{-1} m_{XY}^{\downarrow X}(\{x\}) \cdot m_{Y|X}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}) \\
&= (1-K)^{-1} \cdot 0.9 \cdot (1-K) = 0.9,
\end{aligned}$$

which is in the contradiction with the assumption, because $m_{XY}(\{(x,y),(x,\bar{y})\}) = 0$. □

To simplify the notation, and to make it a bit more lucid, let us denote in the rest of this section $m_{\mathcal{V}|\mathcal{U}} = m_{\mathcal{V}} \ominus m_{\mathcal{V}}^{\downarrow \mathcal{U}\cap\mathcal{V}}$. Moreover, in connection with Definition 2, we will identify situations when BPA $m_{\mathcal{V}|\mathcal{U}\cap\mathcal{V}}$ exists and is, in a way, "adapted" to BPA $m_{\mathcal{U}}$. We will say that $m_{\mathcal{V}|\mathcal{U}\cap\mathcal{V}}$ is *tight* with respect to $m_{\mathcal{U}}$ if for all couples of focal elements **a** and **b** (**a** is a focal element of $m_{\mathcal{U}}$, and **b** is a focal element of $m_{\mathcal{V}|\mathcal{U}\cap\mathcal{V}}$) the following condition holds:

$$\text{for } \forall\, b \in \mathbf{b},\ \exists\, a \in \mathbf{a},\ \text{such that } \{a\} \bowtie \{b\} \neq \emptyset. \tag{7}$$

Proposition 3 *Let two BPAs $m_{\mathcal{U}}, m_{\mathcal{V}}$ are such that $m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$ exists. If $m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$ is tight with respect to $m_{\mathcal{U}}$, then*

$$m_{\mathcal{U}} \rhd_f m_{\mathcal{V}} = m_{\mathcal{U}} \rhd_d m_{\mathcal{V}}.$$

Proof. Recall that for BPA $m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$, the existence of which is assumed,

$$m_{\mathcal{V}} = m_{\mathcal{V}}^{\downarrow \mathcal{V}\cap\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}, \tag{8}$$

and that the *d*-composition is defined

$$m_{\mathcal{U}} \rhd_d m_{\mathcal{V}} = m_{\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}.$$

What are the focal elements of $m_{\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$? Let **a** and **b** be arbitrary focal elements of $m_{\mathcal{U}}$ and $m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$, respectively. Due to Proposition 2, $\left(m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}\right)^{\downarrow \mathcal{V}\cap\mathcal{U}}$ is vacuous, $\mathbf{b}^{\downarrow \mathcal{V}\cap\mathcal{U}} = \Omega_{\mathcal{V}\cap\mathcal{U}}$, and $\mathbf{c} = \mathbf{a} \bowtie \mathbf{b} \neq \emptyset$ is a focal element of $m_{\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$. Therefore, when computing the Dempster's rule of combination $m_{\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$, the corresponding coefficient of conflict (see Eq. 4)

$$K = \sum_{\mathbf{a}\subseteq\Omega_{\mathcal{U}}, \mathbf{b}\subseteq\Omega_{\mathcal{V}} : \mathbf{a}\bowtie\mathbf{b}=\emptyset} m_{\mathcal{U}}(\mathbf{a}) \cdot m_{\mathcal{V}\setminus\mathcal{U}|\mathcal{V}\cap\mathcal{U}}(\mathbf{b}) = 0. \tag{9}$$

The question is whether for a focal element **c** of $m_{\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$ it may happen that $\mathbf{c} = \mathbf{a} \bowtie \mathbf{b}$, and either $\mathbf{a} \neq \mathbf{c}^{\downarrow\mathcal{U}}$, or $\mathbf{b} \neq \mathbf{c}^{\downarrow\mathcal{V}}$. Since $\mathbf{b}^{\downarrow \mathcal{V}\cap\mathcal{U}} = \Omega_{\mathcal{V}\cap\mathcal{U}}$, for $\forall\, a \in \mathbf{a}$, $\exists\, b \in \mathbf{b}$, $\{a\} \bowtie \{b\}$ is a singleton from $\mathbf{c}^{\downarrow\mathcal{U}} \bowtie \mathbf{c}^{\downarrow\mathcal{V}}$ and therefore $\mathbf{a} \subseteq \mathbf{c}^{\downarrow\mathcal{U}}$. Similarly, the assumption that $m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$ is tight with respect to $m_{\mathcal{U}}$ guarantees that $\mathbf{b} \subseteq \mathbf{c}^{\downarrow\mathcal{V}}$. For all $c \in \mathbf{a} \bowtie \mathbf{b}$, $c^{\downarrow\mathcal{U}} \in \mathbf{a}$ from the definition of a join, and therefore $\mathbf{a} \supseteq \mathbf{c}^{\downarrow\mathcal{U}}$. Analogously, $c^{\downarrow\mathcal{V}} \in \mathbf{b}$ yields $\mathbf{b} \supseteq \mathbf{c}^{\downarrow\mathcal{V}}$. So, we have proven that each focal element **c** of $m_{\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$ is created by a single pair of focal elements $\mathbf{c}^{\downarrow\mathcal{U}}$ of $m_{\mathcal{U}}$ and $\mathbf{c}^{\downarrow\mathcal{V}}$ of $m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$. Therefore (using definition from Eq. (3) and Eq. (9)),

$$\begin{aligned}(m_{\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}})(\mathbf{c}) \\ = \sum_{\mathbf{a}\subseteq\Omega_{\mathcal{U}}, \mathbf{b}\subseteq\Omega_{\mathcal{V}} : \mathbf{a}\bowtie\mathbf{b}=\mathbf{c}} m_{\mathcal{U}}(\mathbf{a}) \cdot m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}(\mathbf{b}) = m_{\mathcal{U}}(\mathbf{c}^{\downarrow\mathcal{U}}) \cdot m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}(\mathbf{c}^{\downarrow\mathcal{V}}).\end{aligned} \tag{10}$$

In the same, way we get from Eq. (8) also

$$m_{\mathcal{V}}(\mathbf{c}^{\downarrow\mathcal{V}}) = (m_{\mathcal{V}}^{\downarrow \mathcal{V}\cap\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}})(\mathbf{c}^{\downarrow\mathcal{V}}) = m_{\mathcal{V}}^{\downarrow \mathcal{V}\cap\mathcal{U}}(\mathbf{c}^{\downarrow \mathcal{V}\cap\mathcal{U}}) \cdot m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}(\mathbf{c}^{\downarrow\mathcal{V}}), \tag{11}$$

which gives that, under the given assumptions,

$$m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}(\mathbf{c}^{\downarrow\mathcal{V}}) = \frac{m_{\mathcal{V}}(\mathbf{c}^{\downarrow\mathcal{V}})}{m_{\mathcal{V}}^{\downarrow\mathcal{V}\cap\mathcal{U}}(\mathbf{c}^{\downarrow\mathcal{V}\cap\mathcal{U}})}. \tag{12}$$

Substituting Eq. (12) into Eq. (10), we get exactly the formula from case **(i)** of Definition 3. The fact that case **(ii)** of this definition never creates a focal element of $m_{\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$ follows from the fact that each couple of focal elements $\mathbf{a}$ and $\mathbf{b}$ ($\mathbf{a}$ is a focal element of $m_{\mathcal{U}}$, and $\mathbf{b}$ is a focal element of $m_{\mathcal{V}|\mathcal{U}\cap\mathcal{V}}$) gives rise of a focal element $\mathbf{a} \bowtie \mathbf{b}$ of $m_{\mathcal{U}} \oplus m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$. Thus, whenever case **(ii)** of Definition 3 is used (under the assumptions of this assertion), then it assigns zero. □

Corollary Let two BPAs $m_{\mathcal{U}}, m_{\mathcal{V}}$ are such that $m_{\mathcal{V}|\mathcal{V}\cap\mathcal{U}}$ exists. If $m_{\mathcal{V}}^{\downarrow\mathcal{V}\cap\mathcal{U}}$ is vacuous, or, if $\mathcal{V} \cap \mathcal{U} = \emptyset$, then

$$m_{\mathcal{U}} \triangleright_f m_{\mathcal{V}} = m_{\mathcal{U}} \triangleright_d m_{\mathcal{V}}.$$

Example 2 In this example we show that, generally, d-composition and f-composition of two BPAs may differ from each other. Consider three binary variables X, Y, Z, and m_{XY} and $m_{Z|Y}$ from Table 3.

Table 3: Example when $m_{Z|Y}$ is not tight with respect to m_{XY}.

$\mathbf{a}$	$m_{XY}(\mathbf{a})$
$\{(x,y)\}$	1.00

$\mathbf{a}$	$m_{Z|Y}(\mathbf{a})$
$\{(\bar{y},\bar{z}),(y,z)\}$	1.00

$\mathbf{a}$	$(m_{XY} \triangleright_d m_{Z|Y})(\mathbf{a})$
$\{(x,y,z)\}$	1.00

$\mathbf{a}$	$(m_{XY} \triangleright_f m_{Z|Y})(\mathbf{a})$
$\{(x,y,\bar{z}),(x,y,z)\}$	1.00

Notice that in this example, $m_{Z|Y}$ is not tight with respect to m_{XY} because for $(\bar{y},\bar{z}) \in \{(\bar{y},\bar{z}),(y,z)\}$ there is no element $a \in \{(x,y)\}$ such that $a \bowtie (\bar{y},\bar{z}) \neq \emptyset$.

5 Almond's Captain's Problem

Let us briefly replicate the Captain's problem from the book by Almond (1995). As said in Section 1, this example motivated this research. Namely, when being converted into the form of a compositional model, it defined the same eight-dimensional BPA regardless of the used composition operator.

For the detailed story, we refer the reader to the original book (Almond, 1995), or the paper (Jiroušek et al., 2022) published in this proceedings. The problem concerns the relation of eight variables presented in Table 4. Their mutual relations are in this paper described in a slightly different way than in the cited book. Here we use three

Table 4: Variables for the Captain's decision.

Variable	# states	States	Description
L	2	true, false	Loading is delayed?
F	2	true, false	Weather forecast is foul?
W	2	true, false	Weather in route is foul?
M	2	true, false	Maintenance is done?
R	2	true, false	Ship needs repairs at sea?
D	4	0, 1, 2, 3	Departure delay (in days)
S	4	0, 1, 2, 3	Sailing delay (in days)
A	7	0, 1, 2, 3, 4, 5, 6	Arrival delay (in days)

prior (one-dimensional) belief functions and five low-dimensional (conditional) BPAs (see Table 5). The resulting eight-dimensional BPA is, for example, given by the formula

$$m_{\{L\}} \triangleright m_{\{F\}} \triangleright m_{\{M\}} \triangleright m_{\{D,F,L,M\}} \triangleright m_{\{F,W\}} \triangleright m_{\{M,R\}} \triangleright m_{\{S,W,R\}} \triangleright m_{\{A,D,S\}}. \qquad (13)$$

The ordering of low-dimensional BPAs in Eq. (13) is compatible with the directed graphical model that underlies the Captain's problem in the sense that the conditional for a variable should be composed only after the conditional associated with its parents. Formally, and not using graphs, this property can be formulated that for any compositional model $m_{\mathcal{U}_1} \triangleright m_{\mathcal{U}_2} \triangleright \ldots \triangleright m_{\mathcal{U}_k}$ corresponding to a directed graphical model, for all $j = 2, \ldots, k$, set $\mathcal{U}_j \setminus (\mathcal{U}_1 \cup \ldots \cup \mathcal{U}_{j-1})$ must be singleton, i.e., $|\mathcal{U}_j \setminus (\mathcal{U}_1 \cup \ldots \cup \mathcal{U}_{j-1})| = 1$. Thus, there are other sequences in which the low-dimensional BPAs may be composed without influencing the resulting eight-dimensional one. All of them may be got from Formula (13) by the application of axioms $A3$, $A4$ from Definition 1, and Property 3 from Proposition 1. Another equivalent one is, e.g.,

$$m_{\{F\}} \triangleright m_{\{F,W\}} \triangleright m_{\{M\}} \triangleright m_{\{M,R\}} \triangleright m_{\{S,W,R\}} \triangleright m_{\{L\}} \triangleright m_{\{D,F,L,M\}} \triangleright m_{\{A,D,S\}}. \qquad (14)$$

Not presenting the lists of focal elements of the eight BPAs from Table 5, we cannot show it, but the reader can certainly imagine that verification of the fact that $m_{\{F,W\}}^{\downarrow\{F\}}$ is vacuous (and therefore $m_{\{F,W\}} = m_{\{F,W\}} \ominus m_{\{F,W\}}^{\downarrow\{F\}}$), and that $m_{\{F,W\}}$ is tight with respect to $m_{\{F\}}$ is simple. It is enough to check 2×2 couples of focal elements to verify the latter condition. Thus, it is easy to verify the assumption of Proposition 3, and to show that $m_{\{F\}} \triangleright_f m_{\{F,W\}} = m_{\{F\}} \triangleright_d m_{\{F,W\}}$. The fact that the analogous equality holds for the first three terms of the formula (14) follows directly from Corollary. In a similar way, it is not difficult to show that the eight-dimensional BPA defined by formula (14) does not depend on which operator of composition is used. Nevertheless, one has to realize that it is necessary to show that $m_{\{M,R\}}$ is tight with respect to $m_{\{F\}} \triangleright m_{\{F,W\}} \triangleright m_{\{M\}}$, where the latter BPA (defined as a composition of three low-dimensional BPAs) has 6 focal elements. As a rule, the longer the compositional model, the more focal elements the corresponding BPA has. Thus, Proposition 3 applies to small compositional models, but when one starts considering multidimensional models composed of hundreds of low-dimensional BPAs, its direct application is unrealistic.

Table 5: Low-dimensional BPAs for the Captain's Problem.

Variables	# focal elements	Description
L	3	prior BPA
F	3	prior BPA
M	1	prior BPA: did not perform maintenance before departure
A, D, S	1	rule calculating total delay: A = D + S
D, F, L, M	1	logical function: departure will be delayed one day for each thing wrong
R, S, W	2	noisy logical statement: sailing time increases by one day if something gets wrong
F, W	2	reliability of weather forecast
M, R	9	relationship between maintenance and repairs at sea

6 Summary & Conclusions

The main result of this paper is presented as Proposition 3. It says that, in some situations, the two composition operators yield the same result. It may be interesting because d-composition is generally of much higher computational complexity than f-composition. Nevertheless, Proposition 3 presents only sufficient conditions, not necessary ones. The determination of necessary conditions remains an open problem.

From the exposition, the reader could notice that the analogy between probabilistic and belief-function graphical models is far from being straightforwards. One can always represent any multivariate probability distribution as a directed graphical model (but the directed graphical model may not encode all the conditional independencies in the joint distribution). As shown in Example 1, it is not true for belief functions because there are joint BPAs for which some conditionals do not exist. Similarly, see Remark 1 stepwise composition need not always hold for BPAs. On the other hand, like the d-composition operator, the probabilistic composition operator is not always defined. Surprisingly, f-composition is always defined. It is made possible by the heuristics expressed by case (ii) of Definition 3.

Acknowledgement

This study was financially supported by the Czech Science Foundation under Grant No. 19-06569S, and by the Ronald G. Harper Professorship at the University of Kansas awarded to the third author.

References

R. G. Almond. *Graphical Belief Modeling.* Chapman & Hall, London, UK, 1995.

R. Jiroušek and P. P. Shenoy. Compositional models in valuation-based systems. *International Journal of Approximate Reasoning*, 55(1):277–293, 2014.

R. Jiroušek and P. P. Shenoy. A new definition of entropy of belief functions in the Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 92(1):49–65, 2018.

R. Jiroušek, J. Vejnarová, and M. Daniel. Compositional models for belief functions. In G. de Cooman, J. Vejnarová, and M. Zaffalon, editors, *Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '07)*, pages 243–252, 2007.

R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. Computing the decomposable entropy of graphical belief function models. In *Proc. 12. Workshop on Uncertainty Processing (WUPES 2022)*. MatfyzPress, Praha, Czech Republic, 2022.

G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

G. Shafer. Belief functions and parametric models. *Journal of the Royal Statistical Society, Series B*, 44(3):322–352, 1982.

P. P. Shenoy. Conditional independence in valuation-based systems. *International Journal of Approximate Reasoning*, 10(3):203–234, 1994.

P. P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. In R. D. Shachter, T. Levitt, J. F. Lemmer, and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*, Machine Intelligence and Pattern Recognition Series, vol. 9, pages 169–198. North-Holland, Amsterdam, 1990.

P. Smets. *Un modele mathematico-statistique simulant le processus du diagnostic medical.* PhD thesis, Free University of Brussels, 1978.

Modeling the spread of loanwords in South-East Asia using sailing navigation software and Bayesian networks

František Kratochvíl[1], Václav Kratochvíl[2], George Saad[3], and Jiří Vomlel[4]

[1,3]Department of Asian Studies, Palacký University Olomouc
frantisek.kratochvil,george.saad@upol.cz
[2,4]Institute of Information Theory and Automation, Czech Academy of Sciences
velorex,vomlel@utia.cas.cz

Abstract

A loanword is a word permanently adopted from one language and incorporated into another language without translation. In this paper we study loanwords in the South-East Asia Archipelago, a home to a large number of languages.

Our paper is inspired by the works of Hoffmann et al. (2021) Bayesian methods are applied to probabilistic modeling of family trees representing the history of language families and by Haynie et al. (2014) modelling the diffusion of a special class of loanwords, so called Wanderwörter in languages of Australia, North America and South America. We assume that in the South-East Asia Archipelago Wanderwörter spread along specific maritime trade routes whose geographical characteristics can help unravel the history of Wanderwörter diffusion in the area. For millennia trade was conducted using sailing ships which were constrained by the monsoon system and in certain areas also by strong sea currents. Therefore rather than the geographical distances, the travel times of sailing ships should be considered as a major factor determining the intensity of contacts among cultures.

We use a sailing navigation software to estimate travel times between different ports and show that the estimated travel times correspond well to travel times of a Chinese map of the sea trade routes from the early seventeenth century. We model the spread of loanwords using a probabilistic graphical model - a Bayesian network. We design a novel heuristic Bayesian network structure learning algorithm that learns the structure as a union of spanning trees for graphs of all loanwords in the training dataset. We compare this algorithm with BIC optimal Bayesian networks by measuring how well these models predict the true presence/absence of a loanword. Interestingly, Bayesian networks learned by our heuristic spanning tree based algorithm provide better results than the BIC optimal Bayesian networks.

1 Introduction

This paper examines the loanword distribution in Maritime Southeast Asia, focusing on loanwords from pre-colonial contact languages. The data comes from Blust and Trussel (2013). We know, for example, that the Arabic word *arak* '(type of) alcohol' (and variants thereof) is found in 35 languages of our 124 language sample but are clueless about the borrowing pathway.

In the absence of written records and extensive archaeological and genetic evidence, the distribution of loanwords across a wide area may offer insight into past human migrations, contacts, and trade. The mapping of the loanword distribution offers an opportunity to capture large patterns of human contact that are not as readily detectable by other means. It thus supports the formulation of hypotheses that can later be verified by other disciplines such as history, anthropology, archaeology, and genetics.

Austronesian languages of Maritime Southeast Asia have been the subject of linguistic study for several centuries; their linguistic relationships (proto-languages, branches, etc.) are well described and the Austronesian family belongs to the best reconstructed and documented language families (Greenhill et al., 2008; Blust, 2009; Blust and Trussel, 2013). The languages of Maritime Southeast Asia have accumulated layers of borrowed lexicon from culturally dominant 'donor languages' through trade, conquest, religion and technological development: (i) Old Malay and proto-Malayic (mainly plant and animal terms), (ii) Sanskrit (religion, state), (iii) Tamil (trade), (iv) Arabic (religion, law, trade), (v) Tagalog, and (vi) Chinese (trade). Modern Malay has had the most profound lexical influence in the area, both as a primary source language, and as an intermediary for Sanskrit, Tamil, Indic, Persian, and Arabic words.

Examining the loanwords in Australian and Californian lanugages Haynie et al. (2014) have detected loanword networks using clustering algorithms. Such network is defined by the distribution of so called 'Wanderwörter', which usually make up a small proportion of the total vocabulary of individual languages, and only a minority of loanwords (Haynie et al., 2014). Importantly, Wanderwörter are shared by languages in areas formerly linked by trade (Haynie et al., 2014). The main difference between our work and the work of Haynie and colleagues lies in the nature of the terrain. While the Californian and Australian networks run over land where the distance can be used as the main measure, the Maritime Southeast Asian networks consist of sea routes which are subject to weather patterns (monsoons), sea currents, sea relief, and coastal features. In addition, navigation properties of the various water crafts have to be taken into account when reconstructing the routes.

Our work builds on several previous work and uses a number of publicly available resources. Our primary resource is a large database of loanwords collected from several earlier sources, such as the Austronesian Comparative Dictionary (Blust and Trussel, 2013). The database is available at `http://gogo.utia.cas.cz/loanwords/`.

The paper is structured as follows: Section 2 discusses how travel time was estimated. Section 3 describes the Bayesian network models used. Section 4 elaborates on the experiments used. Finally, section 5 rounds off the paper with some conclusions and suggestions for further research.

2 Estimation of travel time

An important resource for analyzing the historical Maritime Southeast Asian networks is the Selden map, which is the oldest surviving Chinese merchant map of the sea trade routes of East Asia from the early seventeenth century.[1] This map was rediscovered by Robert Batchelor in 2008 in Oxford University's Bodleian Library Batchelor (2013). An important property of the map is that the distances in the map correspond to travel times and therefore the map can serve as a benchmark for estimating the travel times of sailing ships.

In 2021, Perttola (2021) published a method for estimating travel times of sailing ships and compared his estimation with travel times of the Selden map. To some extent, we replicate the work of Perttola (2021) though we implement a different method for computing travel times of sailing ships from the studied historical era. Each sailing ship can be characterized by its polar graph that provides the speed of the ship for different wind angles and wind speed. Based on data presented in Perttola (2021), we estimated the polar graph of the Chinese junk rig, which was a typical sailing ship of that area; see Figure 1 where we present the polar graph of a Chinese junk rig used in our computations.

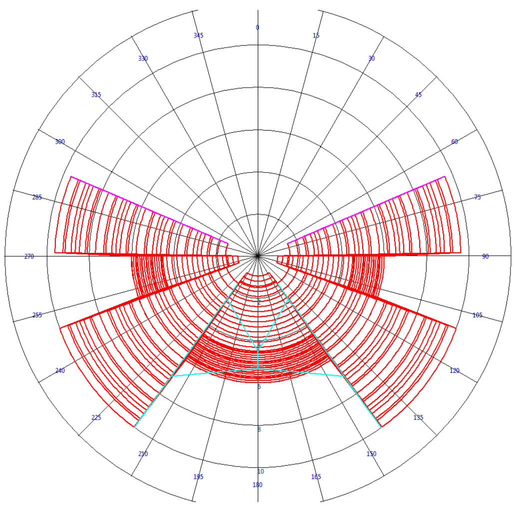


Figure 1: Polar graph of a Chinese junk rig.

In the next step we used a sailing navigation software OpenCPN (2022) and its Weather Routing Plugin (2022) to estimate travel times of the Chinese junk rig specified by its polar graph. The weather routing plugin uses extensive data of winds and see currents from several decades and therefore it can provide reliable estimates of travel times for different seasons.

[1] The map is available at `http://seldenmap.bodleian.ox.ac.uk/` where it is possible to zoom it into a great detail.

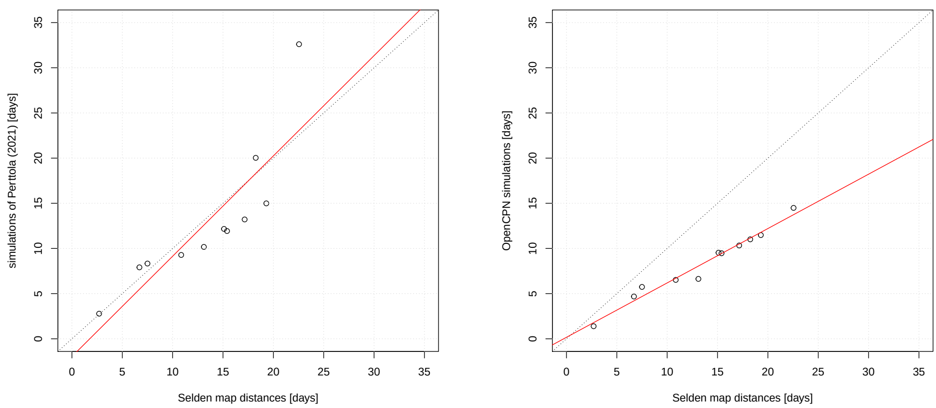
In Figure 2 we present the comparisons of travel times from Selden map and travel estimated by Perttola Perttola (2021) and by us using the sailing navigation software OpenCPN, respectively. We can see that using the sailing navigation software we can estimate the travel times better than Perttola (2021). The Selden map travel times are computed from our estimates as $t_S = 1.66 \cdot t_O - 0.27$ meaning our estimates are systematically lower than true travel times. There can be several reasons for the optimistic estimates - the ships were not always performing as well as expected, the ships may have called at harbours along the route, etc. However, the important result is that the residual standard error of our method, which is 0.6947, is significantly lower than the residual standard error of Perttola's method, which is 4.288.

Perttola (2021)

$t_P = -1.9635 + 1.1103 \cdot t_S$

Our model

$t_O = 0.16469 + 0.60201 \cdot t_S$



Residual standard error: 4.288

Residual standard error: 0.6947

Figure 2: Comparison of travel time estimation methods

In our study we use this method to compute the travel times for all pairs of ports of the area. It serves as a basis for estimation of a Bayesian network (BN) model for the spread of loanwords. In Table 1 we present travel times between main ports of the corresponding language areas. The time is measured in days. Since the Malay speaking region covers a very large area, it is represented by eight different ports, namely by Palembang, Brunei, Banjarmasin, Jakarta, Singkawang, Singapore, Bukit Tengorak, and Samarinda. Similarly, Champa is represented by four different ports, namely by Indrapura, Vijaya, Kauthara, and Panduranga. When computing the distance of Malay (and Champa, respectively) to another language we used always the nearest port. In principle, this could be done also for other languages with several important ports of that language area but we decided to simplify the computations by considering only one port for other languages. We believe this simplification does not have a significant impact on the results presented, especially since many languages correspond to small geographical regions.

Table 1: Travel times (in days) between main ports of the corresponding language areas.

	Acehnese	Aklanon	Balinese	Bikol	Cas. Dumagat	Cebuano	Javanese	Kapampangan	Karo Batak	Makasarese	Manggarai	Maranao	Mongondow	Rejang	Rembong	Sangir	Simalur	Tagalog	Tiruray	Wolio	Iban	Melanau	Champa	Malay
Acehnese		16.22	14.11	18.61	18.96	16.23	11.4	16.17	2.81	15.97	16.3	16.39	20.41	6.58	16.93	17.8	1.46	16.17	16.59	17.43	10.29	10.53	12.15	7.39
Aklanon	16.22		12.28	2.8	3.29	2.59	12.39	2.03	18.7	10.04	10.94	4.02	6	17.11	11.15	4.14	17.34	2.03	3.43	11.38	7.47	6.85	4.93	3.49
Balinese	14.11	12.28		14.63	15.13	11.99	3.47	13.54	12.2	4.09	2.62	11.99	9.63	7.83	3.27	10.78	12.78	13.54	11.86	3.79	7.84	8.14	10.76	2.97
Bikol	18.61	2.8	14.63		0.76	2.67	14.83	3.5	20.95	12.66	13.53	5.88	7.71	19.24	13.17	5.15	19.72	3.5	5.88	12.34	9.84	9.23	6.68	6.1
Cas. Dumagat	18.96	3.29	15.13	0.76		3.2	15.34	3.88	21.18	13.14	14.02	6.25	8.2	19.95	13.65	5.64	20.08	3.88	6.25	12.86	10.38	9.77	6.93	6.59
Cebuano	16.23	2.59	11.99	2.67	3.2		12.35	4.61	18.42	8.36	10.75	3.73	5.59	17.19	12.96	3.66	17.34	4.61	3.1	10.89	7.38	6.77	5.59	3.3
Javanese	11.4	12.39	3.47	14.83	15.34	12.35		12.51	9.58	4.82	5.35	12.61	13.22	4.86	5.86	13.97	10.14	12.51	12.74	6.34	5.33	5.63	7.96	1.95
Kapampangan	16.17	2.03	13.54	3.5	3.88	4.61	12.51		18.42	11.39	12.27	5.97	7.96	17.02	12.48	6.08	17.28	0.5	5.42	12.88	7.53	6.92	4.37	4.78
Karo Batak	2.81	18.7	12.2	20.95	21.18	18.42	9.58	18.42		14.17	14.4	18.73	22.65	5.46	15.05	20.03	2.02	18.42	17.8	15.64	11.57	11.83	13.47	8.7
Makasarese	15.97	10.04	4.09	12.66	13.14	8.36	4.82	11.39	14.17		1.28	9.82	7.66	9.49	1.32	8.58	14.72	11.39	9.7	1.61	9.14	9.46	11.77	3.53
Manggarai	16.3	10.94	2.62	13.53	14.02	10.75	5.35	12.27	14.4	1.28		10.75	8.57	10	0.77	9.51	14.88	12.27	10.61	1.89	9.64	9.97	12.29	4.03
Maranao	16.39	4.02	11.99	5.88	6.25	3.73	12.61	5.97	18.73	9.82	10.75		4.18	17.36	9.85	1.61	17.61	5.97	1	9.16	7.66	7.06	6.44	3.33
Mongondow	20.41	6	9.63	7.71	8.2	5.59	13.22	7.96	22.65	7.66	8.57	4.18		19.78	7.92	2.91	21.53	7.96	4.98	7.1	11.6	11	10.64	4.47
Rejang	6.58	17.11	7.83	19.24	19.95	17.19	4.86	17.02	5.46	9.49	10	17.36	19.78		10.51	18.46	6.06	17.02	17.49	10.98	10.09	10.23	12.56	4.04
Rembong	16.93	11.15	3.27	13.17	13.65	12.96	5.86	12.48	15.05	1.32	0.77	9.85	7.92	10.51		8.5	15.55	12.48	9.66	1.54	10.95	10.47	12.77	4.6
Sangir	17.8	4.14	10.78	5.15	5.64	3.66	13.97	6.08	20.03	8.58	9.51	1.61	2.91	18.46	8.5		18.91	6.08	1.43	7.76	8.98	8.39	7.78	4.6
Simalur	1.46	17.34	12.78	19.72	20.08	17.34	10.14	17.28	2.02	14.72	14.88	17.61	21.53	6.06	15.55	18.91		17.28	17.71	16.2	10.47	10.72	12.33	7.61
Tagalog	16.17	2.03	13.54	3.5	3.88	4.61	12.51	0.5	18.42	11.39	12.27	5.97	7.96	17.02	12.48	6.08	17.28		5.42	12.88	7.53	6.92	4.37	4.78
Tiruray	16.59	3.43	11.86	5.88	6.25	3.1	12.74	5.42	17.8	9.7	10.61	1	4.98	17.49	9.66	1.43	17.71	5.42		8.85	7.77	7.17	6.58	3.44
Wolio	17.43	11.38	3.79	12.34	12.86	10.89	6.34	12.88	15.64	1.61	1.89	9.16	7.1	10.98	1.54	7.76	16.2	12.88	8.85		10.62	10.96	13.25	5.08
Iban	10.29	7.47	7.84	9.84	10.38	7.38	5.33	7.53	11.57	9.14	9.64	7.66	11.6	10.09	10.95	8.98	10.47	7.53	7.77	10.62		0.63	3.29	1.29
Melanau	10.53	6.85	8.14	9.23	9.77	6.77	5.63	6.92	11.83	9.46	9.97	7.06	11	10.23	10.47	8.39	10.72	6.92	7.17	10.96	0.63		2.99	3.89
Champa	12.15	4.93	10.76	6.68	6.93	5.59	7.96	4.37	13.47	11.77	12.29	6.44	10.64	12.56	12.77	7.78	12.33	4.37	6.58	13.25	3.29	2.99		3.19
Malay	7.39	3.49	2.97	6.1	6.59	3.3	1.95	4.78	8.7	3.53	4.03	3.33	4.47	4.04	4.6	4.6	7.61	4.78	3.44	5.08	1.29	3.89	3.19	

The travel times were computed using the Weather Routing Plugin (2022) of the sailing navigation software OpenCPN (2022), The Weather Routing Plugin optimizes ship routes using an isochrone method and predictive grib data or averaged gridded climate data. In this experiment, we performed computations for the date of January 1, which belongs to the period of the winter monsoon. The data from the database of the National Oceanic and Atmospheric Administration (NOAA) represents a 30 year average of winds and currents.[2] The resulting data presented in Table 1 is symmetric since we selected the direction with the shorter travel time for each pair of ports. The background assumption is that the boats can return during summer (when the winds reverse) with comparable travel time as in winter for the reverse direction. Such assumption is justified by the historical evidence showing that ships used to call at the most suitable harbours along their route where they awaited the optimal wind and weather conditions.[3]

3 Bayesian network models

Bayesian networks (Pearl, 1988; Jensen, 2001) are a popular class of probabilistic graphical models (Lauritzen, 1996; Koller and Friedman, 2009), i.e. models that use graphs to describe relations between random variables represented by nodes in the graphs. Bayesian networks model the variables' relations using directed acyclic graphs (DAGs) and their quantitative part is specified by conditional probability tables (CPTs) provided for each variable given its parents in the graph. We believe that Bayesian networks are very suitable for the task of modeling the spread of loanwords since using the graphical part we can use edges to encode frequent contacts between two languages and the quantitative relations between two languages by its nature can be modeled well by probabilistic relations. The state 0 of a model variable represents a particular loanword being absent in the corresponding language while state 1 corresponds to its presence.

To learn the graphical structure of a BN model modeling the spread of loanwords we suggest the following algorithm. Let L be the set of all considered loanwords and $a = 1, 2, 3, \ldots, 17$. The algorithm uses a training dataset which provides information on languages in which each loanword is present.

- Create an empty graph G_0 whose nodes correspond to studied recipient languages.
- For each loanword $\ell \in L$ from training data do:
  - create graph G_ℓ as a copy of graph G_0,
  - add an edge for each language pair containing loanword ℓ to graph G_ℓ. This means that the subgraph generated by languages where the loanword is present is a complete graph.

[2]It is assumed here that the weather systems of the region has been stable during the last three millennia.

[3]A more comprehensive model could include computations for several dates to select the optimal travel time for the journey and return, during the optimal season, should these fall between the monsoon peaks in January and July.

– Evaluate all edges $X - Y$ of graph G_ℓ by the travel time $t_S(X, Y)$ from language X to language Y.
– Find the cheapest spanning tree T_ℓ of graph G_ℓ.

- Make the union of all trees $T_\ell, \ell \in L$ by performing the union on the sets of edges for all studied loanwords. This creates an undirected graph H.
- Assign a weight to each edge computed as a number of trees in $\{T_\ell, \ell \in L\}$ containing this edge.
- Exclude from graph H edges that appear in less than a graphs.
- Direct the edges from the language with the higher number of loanwords.[4] This creates a DAG G that defines the structure of a Bayesian network.

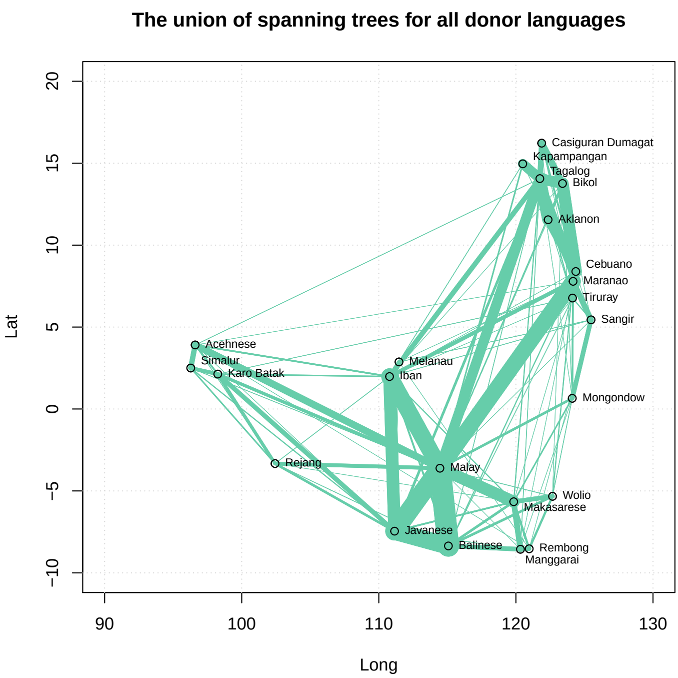


Figure 3: The underlying undirected graph H for $a = 1$.

The database of loanwords is also used to estimate probability values of CPTs which constitute the quantitative part of the learned BN. In Figure 3 we present an example of

[4] Ties are broken randomly.

the underlying undirected graph H with the threshold value $a = 1$ where the width of each edge corresponds to the number of spanning trees the edge is present. The locations of graph nodes correspond to geographical locations of corresponding languages.

In Figure 4 we present a printscreen from our linguistic tool based on a Bayesian network model. Each monitor window represents the marginal probability of a loanword being present in a studied language. The red color bars represent observed evidence for the presence or absence of a loanword while the bars printed green give a prediction probability that this loanword is present in corresponding language.

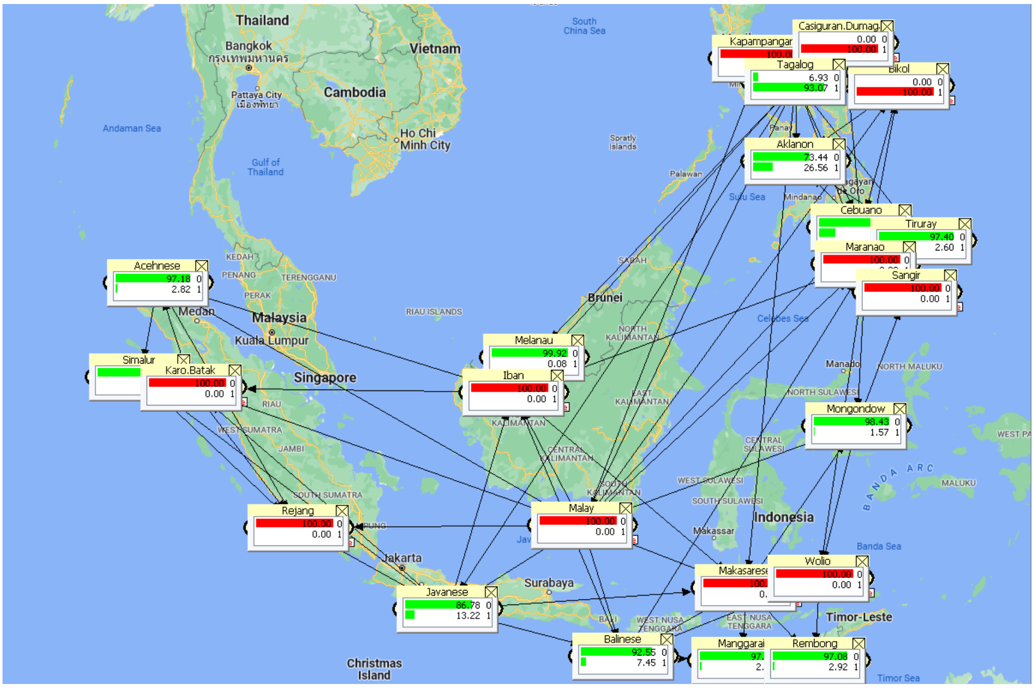


Figure 4: A printscreen from our linguistic tool based on a Bayesian network model.

4 Experiments

Standard methods for structural learning of BNs can also be applied to the studied problem. Therefore we decided to compare our proposed heuristic algorithm with the state-of-the-art learning algorithm which learns a globally optimal BN with respect to data and uses the Bayesian Information Criterion (BIC) as a BN quality criterion. For this purpose we used the Gobnilp method of Cussens and Bartlett (2018). The experiments were performed in the following way:

- The dataset D of 461 loanwords was split to ten subsets $D_1, \ldots, D_n$, all but one containing 46 loanwords and the tenth subset containing 47 loanwords.
- The following steps were repeated for each subset $D_i, i = 1, \ldots, 10$:
  - The set $D \setminus D_i$ was used to learn the BN model using the spanning trees based method (for $a = 1, 2, \ldots, 17$) and the BIC-optimal method.
  - The learned models were tested on loanwords from the set D_i. The presence/absence of this loanword for 11 randomly selected languages was entered into tested models and the conditional probability of the loanword presence in each of the remaining languages was estimated. The predicted state was compared with the true presence/absence of the loanword in D_i.
  - The number of true positives (tp), true negatives (tn), false positives (fp), and false negatives (fn) was computed and added to the results of previous subsets.
- The values of tp, tn, fp, and fn were used to compute the precision, recall, accuracy and balanced accuracy for both methods.

In this way we were able to compare models' predicting ability. In order to estimate the prediction quality limits of the models we repeated the above experiment also for the predictions based on observations of 21 randomly selected languages (instead of 11). The results are summarized in Figure 5. In Figure 6 we give an example of a spanning trees based model with the threshold value $a = 1$ and a BIC optimal model, both learned on the same training dataset.

5 Conclusions and future work

In this paper we studied loanword distribution patterns in Maritime Southeast Asia using the tools based on sailing navigation software and Bayesian networks. The sailing navigation software provided us a deeper insight into travel time between important ports of the studied region. We have shown that this method can provide good travel times estimate corresponding to travel times from a historical map of the sea trade routes in this area.

In the paper we have designed a novel heuristic Bayesian network structure learning algorithm and compared this algorithm with the Gobnilp method that learns BIC optimal Bayesian network structures. Bayesian networks learned by our heuristic spanning tree based algorithm have better prediction quality than the BIC optimal Bayesian networks. This might be attributed to the ability of our heuristic algorithm to exploit additional information provided by travel time distances. However, this needs to be further explored.

Since each language is represented by a Boolean variable, its conditional probability table $P(X|pa(X)$ can be defined as a *noisy-or*. The inhibition probabilities p_i for each $X_i \rightarrow Y$, $X_i \in pa(Y)$ can be learned from collected data. Then the conditional probability of the noisy-or is defined as:

$$P(Y = 0|X_1 = x_1, ..., X_n = x_n) \quad = \quad p_0 \cdot \prod_{i=1}^{n} (p_i)^{x_i} \quad .$$

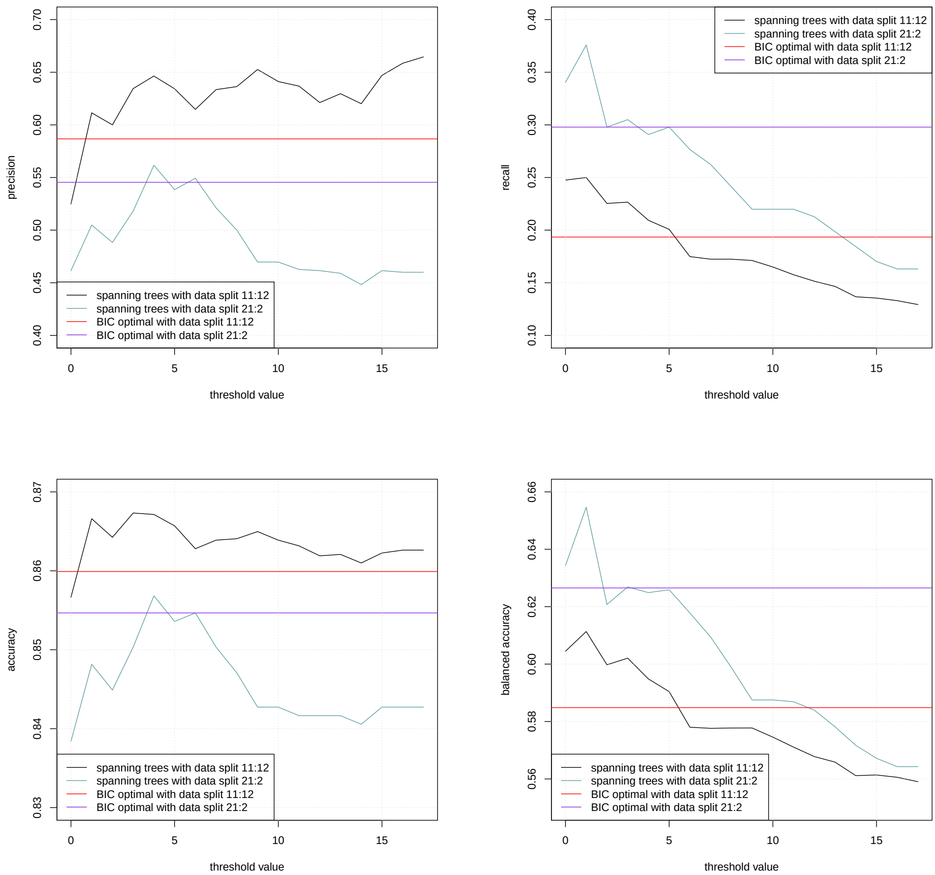


Figure 5: Comparisons of precision, recall, accuracy and balanced accuracy for the spanning trees based models with different threshold value a and the BIC optimal models.

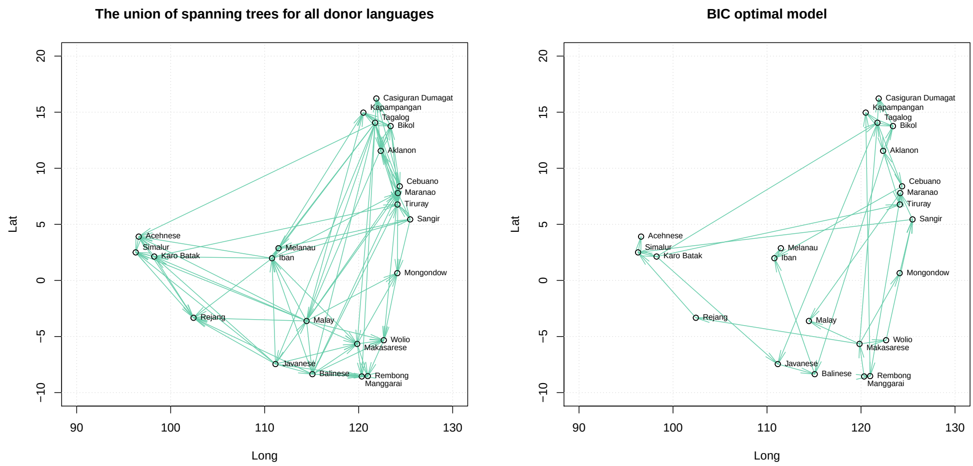


Figure 6: A spanning trees based model with the threshold value $a = 1$ and a BIC optimal model, both learned on the same training dataset.

The noisy-or model has a natural interpretation within this application. If a loanword is present in a language $X_i \in pa(Y)$ it is probable that it is present in language Y as well unless this influence is canceled for some reason. The probability of this influence being canceled is p_i. Naturally, one parent with the loanword being present could be sufficient (therefore the OR relation) but, of course, the more often the loanword is present in $pa(Y)$, the higher the probability of its presence in Y as well. A specialized BN structure learning method might be tailored well for BN models with its CPTs having this particular interpretation.

Also, from the linguistic point of view, it would be interesting to build separate models for each group of loanwords according to their donor language while refining the network of known harbours in each historical period and establishing polar graphs for older ship types than the Chinese junk rig - currently the only type for which a polar graph is available. This would allow an even more refined analysis of the spread of loanwords in different time epochs. We hope to address the above issues in our future work.

Acknowledments

We are grateful to Christian Dumard for recommending the OpenCPN tool. We are also grateful to the authors of this tool who made the weather routing plugin of the OpenCPN tool freely available. This work was supported by the Czech Science Foundation Project Nr. 20-18407S *Verb Class Analysis Accelerator for Low-Resource Languages - Robocorp*. The loanword data was compiled with the support of the ERDF grant *Sinophone Borderlands: Interactions at the Edges* (CZ.02.1.01/0.0/0.0/16_019/00 00791).

References

R. Batchelor. The Selden map rediscovered: A Chinese map of East Asian shipping routes, c.1619. *Imago Mundi*, 65(1):37–63, 2013. doi: 10.1080/03085694.2013.731203.

R. Blust. *The Austronesian languages.* Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University, 2009.

R. Blust and S. Trussel. Australian comparative dictionary: A work in progress. *Oceanic Linguistics*, 52:493—-523, 2013. URL http://dx.doi.org/10.1353/ol.2013.0016.

J. Cussens and M. Bartlett. Gobnilp, 2018. Version version 1.6.3, https://www.cs.york.ac.uk/aig/sw/gobnilp/.

S. D'Epagnier. Weather Routing Pi, 2022. Version 1.13.38, https://opencpn.org/OpenCPN/plugins/weatherroute.html.

S. J. Greenhill, R. Blust, and R. D. Gray. The Austronesian basic vocabulary database: From bioinformatics to lexomics. *Evolutionary Bioinformatics*, 4:271—283, 2008.

H. Haynie, C. Bowerna, P. Epps, J. Hill, and P. McConvell. Wanderwörter in languages of the Americas and Australia. *Ampersand*, 1:1–18, 2014. URL https://doi.org/10.1016/j.amper.2014.10.001.

K. Hoffmann, R. Bouckaert, S. J. Greenhill, and D. Kühnert. Bayesian phylogenetic analysis of linguistic data using beast. *Journal of Language Evolution*, 8:1–75, 2021.

F. V. Jensen. *Bayesian networks and decision graphs.* Springer-Verlag, New York, 2001.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques.* The MIT Press, 2009.

S. L. Lauritzen. *Graphical Models.* Clarendon Press, 1996.

J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

W. Perttola. Digital navigator on the seas of the Selden map of China: Sequential least-cost path analysis using dynamic wind data. *Journal of Archaeological Method and Theory*, 2021. URL https://doi.org/10.1007/s10816-021-09534-6.

The OpenCPN Development Team. OpenCPN, 2022. Version 5.6.0, https://opencpn.org/.

GUHA method and Python Language

Petr Máša[1] and Jan Rauch[1]

[1]Department of Information and Knowledge Engineering, Prague University of Economics and Business
masa@petrmasa.com, rauch@vse.cz

Abstract

Many business users and data owners currently leverages benefits from machine learning algorithms and in many cases, straightforward explain-ability is needed to immediately apply the results of ML methods by simple business decisions. For that reason, association rules based on categorical data is straightforward candidate to use. In modern data science packages (R, Python), arules package is available but it offers only very limited association rules. The promising way is to revitalize GUHA methods. GUHA (General Unary Hypotheses Automaton) is a method of exploratory data analysis developed since 1960s. The method is realized by GUHA procedures. Input of such procedure consists of a definition of a set of relevant patterns and of an analysed data. Output consists of a set of all patterns true in the input data. The GUHA procedure ASSOC deals with patterns, which can be understood as an enhancement of association rules introduced many years later. As so far most widespread implementation of GUHA procedures is LISp-Miner developed in Visual C, we decided to implement key GUHA procedures in modern language, namely in Python. As GUHA offers and LISp-Miner implements more procedures than only association rules, several additional procedures has been implemented. The CF-Miner procedure mines for conditional histograms with a given shape. The SD4ft-Miner procedure mines for interesting couples of conditional association rules. The goal of the paper is to introduce the CleverMiner package as Python package that implements previously mentioned GUHA procedures.

1 Introduction

GUHA is a method of exploratory data analysis developed since 1960s [Hájek et al. (1966)]. GUHA is an abbreviation of *General Unary Hypotheses Automaton*. *General* refers both to a wide field of applications and to generality of results, *Unary* corresponds to the form of analysed data. *Hypotheses* points to hypothesis formation and *Automaton* refers to the use of computers. The method is realized by GUHA procedures. Input of such procedure consists of a definition of a set of relevant patterns and of an analysed data. Output consists of a set of all patterns true in the input data.

The GUHA procedure ASSOC [Hájek and Havránek (1978)] deals with patterns, which can be understood as an enhancement of association rules introduced many years later [Agrawal et al. (1993)]. Since the advent of data mining, the GUHA method is developed as a method of data mining [Hájek et al. (2010)]. Probably the most-used implementation of the ASSOC procedure dealing with enhanced association rules is the procedure 4ft-Miner [Rauch (2013),Rauch and Šimůnek (2017); Turunen and Dolos (2021)]. It is a part of the LISp-Miner system [Šimůnek (2003)]. A comparison of the 4ft-Miner procedure and apriori algorithm is available in [Rauch and Šimůnek (2017)]. A possibility to use expert deduction rues in dealing with domain knowledge in applications of the 4ft-Miner is described in [Rauch (2019)]. The LISp-Miner system includes additional GUHA procedures dealing with various types of pattern. The CF-Miner procedure mines for conditional histograms with a given interesting shape [Rauch and Šimůnek (2019)]. The SD4ft-Miner procedure mines for interesting couples of association rules [Rauch (2013)].

As many machine learning methods exist, very advanced prediction models can be prepared, there is also business urgency on explainability of the model, mainly in situations where model to be used and improper model can spoil some revenue stream. For that reason, managers want to have a model explained to understand it. Advanced models are almost unexplainable as well as classical models lead to incorrect interpretation (e.g. regression with strongly correlated predictors) or can lead in incorrect interpretation (e.g. split in trees is explainable only in context of all splits above this split to root). And for that reason, rule mining on categorical data can be used in this case. Therefore we decided to implement advanced association rules to state-of-the-art language.

Development of the CleverMiner system has been started recently. Its goal is to implement the GUHA procedures in the Python language which is very popular in the Machine Learning and Data Science community. New versions of the 4ft-Miner, CF-Miner and SD4ft-Miner procedures have been implemented. The goal of the paper is to introduce these GUHA procedures of the CleverMiner systems.

The structure of the paper is as follows. Data matrices and Boolean attributes derived from columns of data matrices are introduced in Section 2. These Boolean attributes are used in all described procedures. Then, the data set *Accidents* used in examples of application of particular procedures is briefly described in Section 3. Main features of the CleverMiner system are introduced in Section 4. Examples of applications of the procedures CF-Miner, 4ft-Miner, and SD4ft-Miner are concisely described in Sections 5, 6, and 7 respectively.

2 Data matrix and Boolean attributes

All the described GUHA procedures deal with data matrices in a form shown in Fig. 1. Rows of a data matrix correspond to observed objects, columns correspond to attributes describing particular objects. Each attribute has a finite number of possible values called *categories*. Data matrix $\mathcal{M}$ shown in Fig. 1 has K columns – attributes $A_1, \ldots A_K$.

Patterns particular GUHA procedures deal with concern Boolean attributes derived from columns – attributes of a given data matrix. Basic Boolean attributes are created

$\mathcal{M}$	attributes				Boolean attributes			
	A_1	A_2	$\dots$	A_K	$A_1(1)$	$A_2(3,5)$	$A_1(1) \wedge A_2(3,5)$	$A_1(1) \vee A_2(3,5)$
o_1	1	3	$\dots$	2	1	1	1	1
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
o_n	4	1	$\dots$	6	0	0	0	0

Figure 1: Data matrix $\mathcal{M}$ and examples of Boolean attributes

first. A *basic Boolean attribute* is an expression $A(\alpha)$ where $\alpha \subset \{a_1, \dots a_t\}$ and $\{a_1, \dots a_t\}$ is a set of all categories of the attribute A. The set α is a *coefficient* of the basic Boolean attribute $A(\alpha)$. A *basic Boolean attribute* $A(\alpha)$ *is true in a row* o of $\mathcal{M}$ if $A[o] \in \alpha$. If $A[o] \notin \alpha$, then $A(\alpha)$ is *false in a row* o. Here $A[o]$ denotes a value of the attribute A in the row o. Basic Boolean attributes are also called *literals*.

Each basic Boolean attribute is a Boolean attribute. If φ and ψ are Boolean attributes, then $\varphi \wedge \psi$, $\varphi \vee \psi$ and $\neg\varphi$ are also Boolean attributes. Their values are defined in a usual way. Expressions $A_1(1)$ and $A_2(3,5)$ in Fig. 1 are examples of basic Boolean attributes. Expressions $A_1(1) \wedge A_2(3,5)$ and $A_1(1) \vee A_2(3,5)$ are examples of Boolean attributes.

3 Accidents data set

We use *UK Car Accidents 2005-2015 Data Set* originating from the UK Department for Transport[1]. It concerns traffic accidents in UK in period 2005-2015. There are 1 780 653 accidents. We limit ourselves to accidents concerning one vehicle only to decrease a complexity of analysis. Thus, we use data matrix *Accidents* with 538 989 rows. Nine used attributes are described in Tab. 1. Frequencies of some categories are given in brackets. Note that there are missing values. They were treated by tools of Python, see Section 4. Histogram of the tenth attribute Year meaning a year of the accident is shown in Fig. 2.

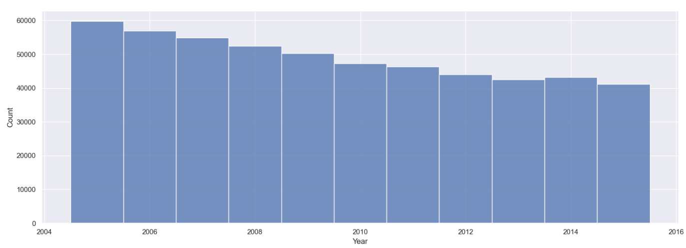


Figure 2: Attribute Year – the numbers of accidents in years 2 005 – 2 015

[1]see `https://www.kaggle.com/silicon99/dft-accident-data`

Table 1: Attributes of the *Accidents* data matrix used in the examples

Attribute	Categories
Vehicle_Type	*Car* (386 346), *Bus_coach_17+* (40 935), ..., 20 categories
Sex	*Male* (365 266), *Female* (137 270) (sex of driver)
Driver_Age_Band	*16 - 20*, *21 - 25* ..., *66 - 75*, *Over 75*
Highway	*E10000016, Kent* (14 777), ... 207 categories
District	*Birmingham* (9 461), *Leeds* (7 365), ... 416 categories
Area	*Urban* (340 200), *Rural* (198 743)
Speed_Limit	*10*, *15*, *20*, ..., *70*
Severity	*Fatal* (10 349), *Serious* (105 754), *Slight* (422 886)
Casualties	*1* (475 314), *2* (48 118), ..., *6* (256), *7 – 68* (230)

4 CleverMiner system – main features

The goal of the CleverMiner system is to bring the GUHA method closer to the Machine Learning and Data Science community. Three GUHA procedures CF-Miner, 4ft-Miner, and SD4ft-Miner are implemented. They are a bit simplified versions of the procedures with the same names implemented in the LISp-Miner system [Šimůnek (2003); Rauch (2013)]. Short descriptions and examples of applications of particular procedures are available in Sections 5, 6, and 7.

CleverMiner is a package for Python available on PyPi (Python Package Index) – standard repository for Python packages. When we want to call GUHA procedure, as input CleverMiner in general requires

df - input pandas dataframe

cedents - definition of sets of relevant Boolean attributes from which patterns to be mined are generated

quantifiers - list of basic quantifiers defining a used quantifier.

An example of input for application of the CF-Miner procedure described in the next section is available in Fig. 4 in Section 5.

CleverMiner procedure prepares dataset into efficient internal binary form (standard dataframe structures are not suitable as many thousands or millions of very similar queries are processed so internal form is much more efficient) and verifies all[2] relevant patterns/rules. Result is in form of Python dictionary (native data structure) that consists of information of task (inputs), processing statistics and list of rules in machine processable form that allow direct or user-friendly printing and machine post-processing list of rules

[2]In fact, not all patterns are verified when optimization is in place but result is the same as if all rules would be verified.

like visualization, filtering etc. To handle missing values, simple Python native method – most frequent simple imputer was used.

```
imputer = SimpleImputer(strategy="most_frequent")
df = pd.DataFrame(imputer.fit_transform(df),columns = df.columns)
```

5 Data mining for increasing histograms

A *conditional histogram* $A/[\chi, TypeHist]$ *of an attribute* A *with categories* $a_1, \ldots, a_t$ *of a type* $TypeHist$ *at a data matrix* $\mathcal{M}$ is a vector $\langle h_1, \ldots, h_t \rangle$ of real numbers. The numbers $h_1, \ldots, h_t$ correspond to heights of particular columns of the histogram. They are computed from a CF-table $CF(A, \mathcal{M}/\chi)$ shown in the left part of Fig. 3. The CF-table $CF(A, \mathcal{M}/\chi)$ is a $2t$-tuple of non-negative integers $\langle n_1, \ldots, n_t, m_1, \ldots, m_t \rangle$. Here, n_i denotes the number of rows o of a data matrix $\mathcal{M}/\chi$ for which $A(o) = a_i$ and m_i denotes the number of rows o of a data matrix $\mathcal{M}$ for which $A(o) = a_i$. We assume $i = 1, \ldots, t$. The histogram $\langle h_1, \ldots, h_t \rangle$ of the type $TypeHist$ computed from the CF-table $CF(A, \mathcal{M}/\chi)$ is denoted by $Hist(CF(A, \mathcal{M}/\chi), TypeHist)$, see the right part of Fig. 3.

$\mathcal{M}$	a_1	$\ldots$	a_t
χ	n_1	$\ldots$	n_t
$True$	m_1	$\ldots$	m_t

CF-table $CF(A, \mathcal{M}/\chi)$

$TypeHist$	$Hist(CF(A, \mathcal{M}/\chi), TypeHist)$			Σ
	h_1	$\ldots$	h_t	
Abs	n_1	$\ldots$	n_t	n
$Proc$	$100 * \frac{n_1}{n}$	$\ldots$	$100 * \frac{n_t}{n}$	100
Cat	$100 * \frac{n_1}{m_1}$	$\ldots$	$100 * \frac{n_t}{m_t}$	–

Heights $\langle h_1, \ldots, h_t \rangle$ of columns of a histogram

Figure 3: CF-table $CF(A, \mathcal{M}/\chi) = \langle n_1, \ldots, n_t, m_1, \ldots, m_t \rangle$

The CF-Miner procedure deals with CF-patterns $\approx_{CF} A/\chi$ where A is a categorical attribute and χ is a Boolean attribute. The symbol $\approx_{CF}$ is a CF-quantifier defining both a type $TypeHist$ of a histogram and a condition of interestingness of a conditional histogram $Hist(CF(A, \mathcal{M}/\chi), TypeHist)$. Here $\mathcal{M}$ is an analysed data matrix.

The trend of accidents in the period 2 005 – 2 015 is decreasing, see Fig. 2. Thus, there is a natural question if there is an interesting segment of accidents with increasing trend in the whole period. We search for a sub-matrix $Accidents/\chi$ of the data matrix $Accidents$ such that χ is a Boolean attribute describing an interesting segment of accidents and that a histogram of the attribute Year at the sub-matrix $Accidents/\chi$ is increasing.

A *step in a histogram* $\langle h_1, \ldots, h_t \rangle$ is each couple $\langle h_i, h_{i+1} \rangle$ where $1 \leq i \leq t-1$. We say that a *step* $\langle h_i, h_{i+1} \rangle$ *is up* if $h_i < h_{i+1}$ and that a *step* $\langle h_i, h_{i+1} \rangle$ *is down* if $h_i > h_{i+1}$. Conditions concerning the numbers of steps up and/or steps down can be used to define CF-quantifiers.

There are three types of histograms $\langle h_1, \ldots, h_t \rangle$ given by the value of $TypeHist$, see the right part of Fig. 3. We are interested in histograms of the attribute Year

at the sub-matrix *Accidents*/χ where the heights of columns correspond to number of accidents in particular years, formally $Hist(CF(\mathsf{Year}, Accidents/\chi), Abs)$. The attribute Year has 11 categories 2 005, ..., 2 015. Thus, there are 10 steps in each histogram $Hist(CF(\mathsf{Year}, Accidents/\chi), Abs)$. Such histogram is increasing if there are 10 steps up. In addition, we will require that there are at least 5 000 accidents satisfying χ. This requirement can be expressed by the CF-quantifier
`quantifiers= 'S_Up': 10, 'Base': 5000`, see Fig. 4 below.

There are rich possibilities of a definition of a set $\mathcal{R}_\chi$ of relevant conditions χ corresponding to relevant Boolean attributes. The relevant Boolean attributes are called *cedents*, it comes from *antecedents* and *succedents* used in association rules.

The core of a definition of a set of relevant cedents is a list of definitions of sets of relevant literals. Definition L of a set $\mathcal{L}(L)$ of relevant literals is in a form:
{`'name': '`*Att*`', 'type': '`*Tp*`', 'minlen':` L_{min}`, 'maxlen':` L_{max}}.
Here *Att* is a name of attribute, *Tp* is a *type of coefficients*, L_{min} and L_{max} are natural numbers satisfying $1 \leq L_{min} \leq L_{max}$, and *Category* is a category of the attribute *Att*.

Recall that a literal is an expression $A(\alpha)$ where A is an attribute and α is a coefficient of $A(\alpha)$ – a subset of a set of categories of A. The number of categories in α is called a *length of coefficient* α. The numbers L_{min} and L_{max} define a *minimal length of coefficient* and a *maximal length of coefficient.*

There are five types of coefficients: `subset`, `one`, `seq`, `lcut`, and `rcut`. We use the types `subset`, `seq`, and `lcut` only. The attributes Vehicle_Type and Driver_Age_Band are used to present examples of coefficients. Each coefficient α of a Boolean attribute $A(\alpha)$ is of type `subset`. Thus, Vehicle_Type(*Car*) is an example of literal with a coefficient of type `subset` of length 1. A basic Boolean attribute Vehicle_Type(*Car*,*Tram*) is an example of literal with a coefficient of type `subset` of length 2.

The remaining types of coefficients are suitable for ordinal attributes. An attribute is *ordinal* if there is a meaningful ranking of its categories. The attribute Driver_Age_Band with categories *16 - 20*, *21 - 25* ..., *66 - 75* is ordinal. Let us further assume that an ordinal attribute A has t categories ranked as follows: $a_1, a_2, \ldots, a_{t-1}, a_t$.

The type `seq` is called *sequence.* A coefficient α of a literal $A(\alpha)$ is of type *sequence* if it holds $A(\alpha) = A(a_u, \ldots, a_v)$ where $1 \leq u \leq v \leq t$. Note that $v - u + 1$ is a length of the coefficient of $A(a_u, a_{u+1}, \ldots, a_{v-1}, a_v)$. A coefficient of a literal Driver_Age_Band(*16 - 20*, *21 - 25*,*26 - 30*) is of the type sequence. A coefficient of a literal Driver_Age_Band(*16 - 20*,*26 - 30*) is not of the type sequence.

If *type* = `lcut` i.e. *left cuts*, then the expression $\mathcal{L}$ defines a set $\mathcal{S}(\mathcal{L})$ of all literals $A(\alpha) = A(a_1, a_2, \ldots, a_{v-1}, a_v)$ where it holds $L_{min} \leq v \leq L_{max}$. Note that v is a length of the coefficient of $A(a_1, a_2, \ldots, a_{v-1}, a_v)$. Driver_Age_Band(*16 - 20*, *21 - 25*) is an example of literal with a coefficient of type *lcut* and length two.

Recall that $\mathcal{L}(L)$ denotes a set of literals defined by L. A definition of a set of relevant cedents is an expression
`'attributes':[`L_1`,...,`L_u`],'minlen':`C_{min}`,'maxlen':`C_{max}`,'type':'`TC`'`.
Here C_{min} and C_{max} are positive integers satisfying $C_{min} \leq C_{max}$. TC denotes a *type of cedent.* It holds either TC `= con` or TC `= dis`.

If TC `= con`, then the definition of a set of relevant cedents s defines a set of relevant cedents as a set of conjunctions $A_{i_1}(\alpha_{i_1}) \wedge \cdots \wedge A_{i_1}(\alpha_{i_u})$ such that $i_1 < \cdots < i_u$, $A_{i_1}(\alpha_{i_j}) \in \mathcal{L}(L_{i_j})$ for $j = 1, \ldots, u$ and $C_{min} \leq u \leq C_{max}$. If TC = `dis`, then a set of relevant cedents is defined as a set of disjunctions $A_{i_1}(\alpha_{i_1}) \vee \cdots \vee A_{i_1}(\alpha_{i_u})$ such that $i_1 < \cdots < i_u$, $A_{i_1}(\alpha_{i_j}) \in \mathcal{L}(L_{i_j})$ for $j = 1, \ldots, u$ and $C_{min} \leq u \leq C_{max}$.

A definition of the set of $\mathcal{R}_\chi$ of relevant conditions χ used to search interesting segments of accidents with increasing trend in the whole period 2 005 – 2015 is shown in Tab. 2. A part of this definition written in Python is available in Fig. 4.

Table 2: Definition of the set of relevant $\mathcal{R}_\chi$ of relevant conditions χ

Cedent: `cond` `maxlen`: 4		`minlen`: 0 `type`: `con`	
`name`	`type`	`minlen`	`maxlen`
Vehicle_Type	subset	1	1
Sex	subset	1	1
Driver_Age_Band	seq	1	3
Highway	subset	1	1
District	subset	1	1
Area	subset	1	1
Speed_Limit	subset	1	1
Severity	subset	1	1
Casualties	seq	1	3

A run of the CF-Miner procedure with the above described parameters resulted in 49 segments of accidents with increasing trends. 8 838 conditional histograms were verified in 14 Seconds. The structure of resulting segments of accidents is as follows.

- A definition χ of each of 49 output segments is a conjunction of at least two literals, one of them being Speed_Limit(*20*).
- There are 6 output segments defined by a conjunction containing one of literals Driver_Age_Band(*26 - 35,36 - 45,46 - 55*), Driver_Age_Band(*21 - 25,26 - 35,36 - 45*), Driver_Age_Band(*26 - 35,36 - 45*). Some of these definitions contain also literals Area(*Urban*) or Severity(*Slight*)
- There are 4 output segments defined by a conjunction containing a literal Sex(*Male*). Some of these definitions contain also literals Area(*Urban*) or Severity(*Slight*)
- There are 3 output segments defined by a conjunction containing a literal Vehicle_Type(*Car*). Some of these definitions contain also literals Area(*Urban*) or Severity(*Slight*).

The largest segment consisting of 8 980 accidents is defined by the conjunction Casualties(*1*) $\wedge$ Speed_Limit(*20*). Thus, it corresponds to the histogram $Hist(CF($Year$, Accidents/($Casualties(*1*) $\wedge$ Speed_Limit(*20*)$), Abs)$ shown in Fig. 5.

```
clm = cleverminer(df=df,target='Year',proc='CFMiner',
quantifiers= {'S_Up':10, 'Base':5000},
cond ={
  'attributes':[
    {'name': 'Vehicle_Type', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
    {'name': 'Sex', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
    {'name': 'Driver_Age_Band', 'type': 'seq', 'minlen': 1, 'maxlen': 3},
    {'name': 'Highway', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
    {'name': 'District', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
    {'name': 'Area', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
    {'name': 'Speed_limit', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
    {'name': 'Severity', 'type': 'subset', 'minlen': 1, 'maxlen': 1},
    {'name': 'Casualties', 'type': 'seq', 'minlen': 1, 'maxlen': 3},
  ], 'minlen':1, 'maxlen':4, 'type':'con'}
)
print(clm.result)
clm.print_rulelist()
```

Figure 4: Input of a run of the CF-Miner procedure described in Section 5

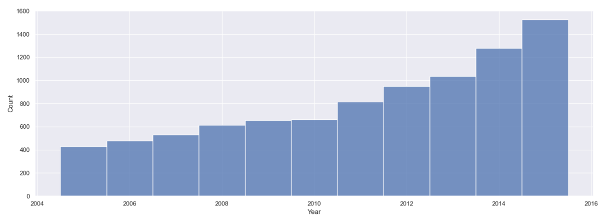


Figure 5: Histogram $Hist(CF(\textsf{Year}, Accidents/(\textsf{Casualties}(1) \wedge \textsf{Speed\_Limit}(20)), Abs)$

6 Association rules – fatal or serious accidents

The attribute Severity has categories *Fatal*, *Serious*, and *Slight*. It is natural to ask which circumstances are related to fatal or serious accidents. We ask which circumstances lead to at least 150 per cent higher relative frequency of fatal or serious accidents than it the whole set of accidents. We use the 4ft-Miner procedure.

The 4ft-Miner deals with association rules $\varphi \approx_{4ft} \psi$ and with conditional association rules $\varphi \approx_{4ft} \psi/\chi$. The symbol $\approx_{4ft}$ is a *4ft-quantifier*. It corresponds to a condition concerning contingency tables of φ and ψ in a data matrix $\mathcal{M}$, see Tab. 3. Such contingency tables are called *4ft-tables* and denoted by 4ft(φ,ψ, $\mathcal{M}$). Here a is the number of rows of a data matrix $\mathcal{M}$ satisfying both φ and ψ, b is the number of objects satisfying, etc. We write $4\text{ft}(\varphi,\ \psi, \mathcal{M}) = \langle a, b, c, d\rangle$.

Table 3: 4ft-table 4ft(φ,ψ, $\mathcal{M}$) of φ and ψ in $\mathcal{M}$

$\mathcal{M}$	ψ	$\neg\psi$
φ	a	b
$\neg\varphi$	c	d

An association rule $\varphi \approx_{4ft} \psi$ is true in a data matrix $\mathcal{M}$ if the condition corresponding to the 4ft-quantifier $\approx_{4ft}$ is satisfied for a 4ft-table 4ft(φ, ψ, $\mathcal{M}$). A conditional association rules $\varphi \approx_{4ft} \psi/\chi$ is true in $\mathcal{M}$ if the rule $\varphi \approx_{4ft} \psi$ is true in a sub-matrix $\mathcal{M}/\chi$.

There are several basic 4ft-quantifiers. An applied 4ft-quantifier $\approx_{4ft}$ is then a conjunction of used basic 4ft-quantifiers. We use two basic 4ft-quantifiers – `'Base':` with parameter B defining a condition $a \geq B$ and `'aad':` with parameter p defining a condition $\frac{a}{a+b} \geq (1+p)\frac{a+c}{a+b+c+d}$. Note that if $\frac{a}{a+b} \geq (1+p)\frac{a+c}{a+b+c+d}$ is satisfied for a 4ft-table $4ft(\varphi, \psi, \mathcal{M}) = \langle a, b, c, d\rangle$, then we can say that a relative frequency of rows of $\mathcal{M}$ satisfying ψ among rows satisfying φ is at least $100p$ per cent higher than a relative frequency of rows satisfying ψ in the whole data matrix $\mathcal{M}$. Let us also note that this corresponds to the fact that a lift of the corresponding association rule is at least $p+1$.

We use a run of the 4ft-Miner mining for true association rules $\varphi \approx_{4ft} \psi$ where the 4ft-quantifier is defined by the expression `quantifiers= 'aad': 1.0, 'Base':5000`, i.e. $p = 1.0$ and $B = 5000$. The set $\mathcal{R}_\varphi$ of relevant antecedents is defined according to Tab. 2; however, the attributes Severity and Casualties are not used. The set $\mathcal{R}_\psi$ of relevant succedents is defined as a set of left cuts with length 1 – 2. This means that $\mathcal{R}_\psi$ = {Severity(*Fatal*), Severity(*Fatal*,*Serious*)}.

A run of the 4ft-Miner with these parameters resulted in 10 rules. 215 482 rules were verified in 19 seconds. The strongest rule (i.e. a rule with the highest possible p such that $\frac{a}{a+b} \geq (1+p)\frac{a+c}{a+b+c+d}$ is satisfied) is
Area(*Rural*) $\wedge$ Sex(*Male*) $\wedge$ Vehicle_Type(*Motorcycle over 500cc*) $\approx_{4ft}$ Severity(*Fatal*, *Serious*).
A 4ft-table of this rule follows:

Accidents	succedent	$\neg$ succedent
antecedent	6 227	5 639
$\neg$ antecedent	109 876	417 247

This means that a relative frequency of accidents satisfying Severity(*Fatal*, *Serious*) in the whole data matrix *Accidents* is $\frac{6227+109876}{6227+5639+109876+417247} = 0.215$. A relative frequency of accidents satisfying Severity(*Fatal*, *Serious*) among the rows satisfying antecedent is $\frac{6227}{6227+5639} = 0.524$, which is 144 percent higher than in the entire data matrix.

The structure of the resulting set of 10 rules can be described in the following way:

- A succedent of each rule is Severity(*Fatal*, *Serious*).
- All antecedents involve Vehicle_Type(*Motorcycle over 500cc*).

- Antecedents of 6 rules involve the attribute Driver_Age_Band with various coefficients.
- Antecedents of 5 rules involve the literal Sex(*Male*)
- Antecedents of 2 rules involve the literal Area(*Rural*).

7 Comparing male and female drivers

Another interesting question is about the differences between men and women regarding the differences in the probability of fatal or serious accidents. We are interested in circumstance under which there is at least 1.5 times higher probability of fatal or serious accidents for men/women than for women/men.

The SD4ft-Miner mines for SD4ft-patterns $\alpha \bowtie \beta : \varphi \approx_{SD4ft} \psi$. The symbol $\approx_{SD4ft}$ is an SD4ft-quantifier. It corresponds to a condition concerning 4ft-tables 4ft(φ,ψ, $\mathcal{M}/\alpha$) and 4ft(φ,ψ, $\mathcal{M}/\beta$), see Fig. 6. An SD4ft-pattern $\alpha \bowtie \beta : \varphi \approx_{SD4ft} \psi$ is true in a data

$\mathcal{M}/\alpha$	ψ	$\neg\psi$
φ	a_α	b_α
$\neg\varphi$	c_α	d_α

4ft(φ,ψ, $\mathcal{M}/\alpha$)

$\mathcal{M}/\beta$	ψ	$\neg\psi$
φ	a_β	b_β
$\neg\varphi$	c_β	d_β

4ft(φ,ψ, $\mathcal{M}/\beta$)

Figure 6: 4ft-tables 4ft(φ,ψ, $\mathcal{M}/\alpha$) and 4ft(φ,ψ, $\mathcal{M}/\beta$)

matrix $\mathcal{M}$ if the condition corresponding to the SD4ft-quantifier $\approx_{SD4ft}$ is satisfied for 4ft-tables 4ft(φ,ψ, $\mathcal{M}/\alpha$) and 4ft(φ,ψ, $\mathcal{M}/\beta$).

There are several basic SD4ft-quantifiers. An applied SD4ft-quantifier $\approx_{SD4ft}$ is a conjunction of used basic SD4ft-quantifiers. We use three basic 4ft-quantifiers – `'Base1':` with parameter B_1 defining a condition $a_\alpha \geq B1$, `'Base2':` with parameter B_2 defining a condition $a_\beta \geq B2$ and `'Ratioconf':` with parameter p defining a condition $\frac{\frac{a_\alpha}{a_\alpha+b_\alpha}}{\frac{a_\beta}{a_\beta+b_\beta}} \geq p$.

Note that if $\frac{\frac{a_\alpha}{a_\alpha+b_\alpha}}{\frac{a_\beta}{a_\beta+b_\beta}} \geq p$ is satisfied for 4ft-tables 4ft(φ,ψ, $\mathcal{M}/\alpha$) and 4ft(φ,ψ, $\mathcal{M}/\beta$), then we can say that a relative frequency of ψ among rows satisfying φ is at least p-times higher for a data sub-matrix $\mathcal{M}/\alpha$ than for a a data sub-matrix $\mathcal{M}/\beta$.

We use a run of the SD4ft-Miner mining for true SD4ft-patterns $\alpha \bowtie \beta : \varphi \approx_{SD4ft} \psi$ where the SD4ft-quantifier is defined by the expression
`quantifiers= 'Base1':4000, 'Base2':4000, 'Ratioconf' :  1.5`. The set $\mathcal{R}_\varphi$ of relevant antecedents is defined according to Tab. 2; however, the attributes Sex, Severity, and Casualties are not used and length of antecedent is 1 – 2. The set $\mathcal{R}_\psi$ of relevant succedents is defined as a set of left cuts with length 1 – 2. This means that $\mathcal{R}_\psi$ = {Severity(*Fatal*), Severity(*Fatal,Serious*)}. The sets $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$ defining sub-matrices $\mathcal{M}/\alpha$ and $\mathcal{M}/\beta$ are defined such that T$\mathcal{R}_\alpha = \mathcal{R}_\beta$ = {Sex(*Female*), {Sex(*Male*)} (subsets with length 1–1).

A run of the SD4ft-Miner procedure with these parameters resulted in six SD4ft-patterns. During the run, there was 2 388 106 pattern verifications. The strongest patterns (a pattern rule with the highest possible p such that $\frac{\frac{a_\alpha}{a_\alpha+b_\alpha}}{\frac{a_\beta}{a_\beta+b_\beta}} \geq p$ is satisfied) is

$\mathsf{Sex}(\mathit{Male}) \bowtie \mathsf{Sex}(\mathit{Female})$: $\mathsf{Area}(\mathit{Rural}) \wedge \mathsf{Speed\_limit}(\mathit{60}) \approx_{SD4ft}$ Severity$(\mathit{Fatal,Serious})$
Corresponding 4ft-tables follow ($(\mathit{Fatal,Serious})$ used instead of Severity$(\mathit{Fatal,Serious})$:

Accidents/Sex(*Male*)	(*Fatal,Serious*)	¬ (*Fatal,Serious*)
Area(*Rural*) ∧ Speed_limit(*60*)	18 487	50 292
¬ (Area(*Rural*) ∧ Speed_limit(*60*))	73 113	259 827
Accidents/Sex(*Female*)	(*Fatal,Serious*)	¬ (*Fatal,Serious*)
Area(*Rural*) ∧ Speed_limit(*60*)	4 391	24 096
¬ (Area(*Rural*) ∧ Speed_limit(*60*))	20 112	88 671

This means that a relative frequency of accidents satisfying Severity(*Fatal*, *Serious*) among accidents satisfying Area(*Rural*) ∧ Speed_limit(*60*) is for male drivers $\frac{18487}{18487+50292} = 0.269$ which is 1.74 times higher than for female drivers $\frac{4391}{4391+24096} = 0.154$.

The structure of the resulting set of six rules can be described in the following way: There are SD4ft-patterns (we denote $p = \frac{a_\alpha}{a_\alpha+b_\alpha} / \frac{a_\beta}{a_\beta+b_\beta}$)
Sex(*Male*) ⋈ Sex(*Female*): Speed_limit(*60*) $\approx_{SD4ft}$ Severity(*Fatal,Serious*) ($p = 1.736$)
Sex(*Male*) ⋈ Sex(*Female*): Area(*Rural*) $\approx_{SD4ft}$ Severity(*Fatal,Serious*) ($p = 1.584$)
Sex(*Male*) ⋈ Sex(*Female*):
Speed_limit(*60*) ∧ Area(*Rural*) $\approx_{SD4ft}$ Severity(*Fatal,Serious*) ($p = 1.744$).
Three SD4ft-pattterns are in a form
Sex(*Male*) ⋈ Sex(*Female*):
Area(*Rural*) ∧ Driver_Age_Band(ω) $\approx_{SD4ft}$ Severity(*Fatal,Serious*),
where ω is one of coefficients (*16 - 20*, *21 - 25*, *26 - 35*), (*21 - 25*, *26 - 35*, *36 - 45*), (*26 - 35*, *36 - 45*,*46 - 55*).

8 Conclusions

We have shortly introduced the GUHA method and informally described applications of three GUHA procedures implemented in the Python language. All of them are a bit simplified versions of the GUHA procedures with the same names implemented in the LISp-Miner system. The 4ft-Miner procedure is an enhanced procedure ASSOC invented in [Hájek and Havránek (1978)]. It is shown in [Rauch and Šimůnek (2017)] that the 4ft-Miner implemented in the LISp-Miner has important advantages when comparing with the apriori algorithm. It this paper, the advantages of the 4ft-Miner are demonstrated by applications of coefficients of types sequences and left cuts. The procedure CF-Miner and SD4ft-Miner remarkably enhance analytical possibilities of the 4ft-Miner procedure.

Further work is related to teaching with both systems, to general research and to widespread generalized association rules to data science community. CleverMiner allows to use many enhancements for users with standard data science skills (like Python experience) without need to contact authors for changes/implementation of new features as results are in machine-processable form. Additional Research involves also dealing with domain knowledge, see e.g. [Rauch (2019); Rauch and Šimůnek (2014)].

References

R. Agrawal, T. Imielinski, and A. N. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, May 26-28, 1993*, pages 207–216, 1993. URL `https://doi.org/10.1145/170035.170072`.

P. Hájek and T. Havránek. The GUHA method – its aims and techniques (twenty-four questions and answers). *International Journal of Man-Machine Studies*, 10(1):3–22, 1978. ISSN 0020-7373. doi: 10.1016/S0020-7373(78)80031-5.

P. Hájek, I. Havel, and M. Chytil. The GUHA method of automatic hypotheses determination. *Computing*, 1(4):293–308, 1966. doi: 10.1007/BF02345483.

P. Hájek, M. Holeňa, and J. Rauch. The GUHA method and its meaning for data mining. *J. Comput. Syst. Sci.*, 76(1):34–48, 2010. URL `https://doi.org/10.1016/j.jcss.2009.05.004`.

J. Rauch. *Observational Calculi and Association Rules*, volume 469 of *Studies in Computational Intelligence*. Springer, 2013. ISBN 978-3-642-11736-7. doi: 10.1007/978-3-642-11737-4.

J. Rauch. Expert deduction rules in data mining with association rules: a case study. *Knowledge and Information Systems*, 59(1):167–195, 2019. ISSN 0219-1377.

J. Rauch and M. Šimůnek. Learning association rules from data through domain knowledge and automation. In A. Bikakis, P. Fodor, and D. Roman, editors, *Rules on the Web. From Theory to Applications - 8th International Symposium, RuleML 2014*, volume 8620 of *Lecture Notes in Computer Science*, pages 266–280. Springer, 2014. doi: 10.1007/978-3-319-09870-8_20.

J. Rauch and M. Šimůnek. Apriori and GUHA - comparing two approaches to data mining with association rules. *Intell. Data Anal.*, 21(4):981–1013, 2017. doi: 10.3233/IDA-160069. URL `https://doi.org/10.3233/IDA-160069`.

J. Rauch and M. Šimůnek. Data mining with histograms and domain knowledge - case studies and considerations. *Fundam. Inform.*, 166(4):349–378, 2019. doi: 10.3233/FI-2019-1805.

E. Turunen and K. Dolos. Revealing drivers natural behavior – a GUHA data mining approach. *Mathematics*, 9(15), 2021. ISSN 2227-7390. doi: 10.3390/math9151818.

M. Šimůnek. Academic KDD project LISp-Miner. In A. Abraham, K. Franke, and M. Köppen, editors, *Intelligent Systems Design and Applications*, pages 263–272, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. ISBN 978-3-540-44999-7.

Inversion of Bayesian Networks

Jesse van Oostrum[1], Peter van Hintum[2], and Nihat Ay[1, 3, 4]

[1]Hamburg University of Technology, Hamburg, Germany
jesse.van@tuhh.de
[2]University of Oxford, Oxford, UK
[3]Leipzig University, Leipzig, Germany
[4]Santa Fe Institute, Santa Fe, USA

Abstract

Variational autoencoders and Helmholtz machines use a recognition network (encoder) to approximate the posterior distribution of a generative model (decoder). In this paper we study the necessary and sufficient properties of a recognition network so that it can model the true posterior distribution exactly. These results are derived in the general context of probabilitic graphical modelling / Bayesian networks, for which the network encodes a set of conditional independency relations. We derive both global conditions, in terms of d-separation and local conditions for the recognition network to have the desired qualities. It turns out that for the local conditions the property *perfectness* (for every node, all parents are connected) plays an important role.

Introduction

A generative model is a set of probability distributions that models the distribution of observed and latent variables. Generative models are used in many machine learning applications. One is often interested in performing inference of the latent variable given an observation, i.e. obtain the posterior distribution. For complex generative models it is often hard to calculate the posterior distribution analytically. The field of variational Bayesian inference Wainwright et al. (2008) studies different ways of approximating the true posterior. Useful machine learning architectures for this are the variational autoencoder (VAE) Kingma and Welling (2013) and Helmholtz machine Dayan et al. (1995). In its most general form these consist of a Bayesian network that is used to model the generative distribution. A second network, called the recognition model, is used to model the posterior distribution. Both these networks have the same set of nodes, namely the union of the observed and latent variables. However, in the generative network the arrows point from the latent to the observed nodes but in the recognition network it is the other way around. The recognition network is therefore in some sense an inversion of the generative network. In many applications, one simply flips the direction of the edges of the generative network to obtain the recognition network. However, as the simple example in Figure 1 shows, this does not guarantee that the recognition model is actually able to model the true posterior distribution of the generative model. In this paper, we study the necessary and sufficient properties of the recognition network such that we do have this guaranty.

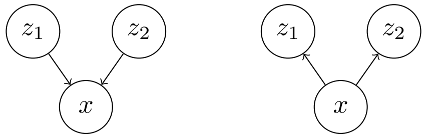


Figure 1: Pair of DAGs G (left) G' (right) where G' is obtained by flipping the direction of the edges in G. The variables z_1, z_2 represent the latent variables and x the observed variable. The distribution p such that z_1, z_2 are Bernoulli(0.5) and $x = z_1 + z_2 \mod 2$ can be modelled by G, but the the conditional distribution $p_{z_1,z_2|c}$ cannot be modelled by G'.

In practice, one often puts further restrictions on the probability distributions the networks can model by for example letting the distribution of an individual node be Gaussian, with the mean and variance being a function of the values of the parent nodes. These type of restrictions fall outside of the scope of this paper.

Markov equivalence is a property of a pair of Bayesian networks that indicates that they encode the same set of conditional independence statements Verma and Pearl (1990); Flesch and Lucas (2007). A generalisation of this, that we will call *Markov inclusion*, is when the set of conditional independence statements encoded in one graph is subset of the other. We will see in Proposition 4 that the results in this paper can also be viewed as describing under which conditions one Bayesian network is Markov inclusive of another.

Notation

Graph theory

For a comprehensive overview of the theory and terminology of probabilistic graphical models, we refer to Lauritzen (1996); Cowell et al. (1999). Let $G = (N, E)$ be a directed acyclic graph (DAG), that we always assume to be connected. The set of parents, children, descendants, and non-descendants of a node $s \in N$ are denoted $\mathrm{pa}(s)$, $\mathrm{ch}(s), \mathrm{des}(s), \mathrm{nd}(s)$ respectively. G is called *perfect* if for all s, the set $\mathrm{pa}(s)$ is complete. For a subset $A \subset N$, the vertex induced subgraph of G is denoted $G[A]$. We let $\mathrm{Leaves}(G) = \{s \in N : \mathrm{ch}(s) = \emptyset\}$ be the set of nodes without children, and $\mathrm{Roots}(G) = \{s \in N : \mathrm{pa}(s) = \emptyset\}$ be the set of nodes without parents. If A is d-separated from B by S we will write $A \perp_d B \mid S$. For $e = (s, t) \in E$, let $e^* = (t, s)$, $E^* = \{e^* : e \in E\}$, $G^* = (N, E^*)$ the graph G with its edges reversed, $G^\sim = (N, E \cup E^*)$, the undirected version of G, and G^{M} the moral graph of G. A *trail* γ in G is a sequence of vertices that form a path in $G^{\sim}$. For two trails $\gamma_1 = (s, ..., t), \gamma_2 = (t, ..., u)$, we write $\gamma_1 + \gamma_2 = (s, ..., t, ...u)$ for the concatenation of the two trails. For a trail $\gamma = (u_1, ..., u_n)$, we write $\gamma_{s,t} = (u_i, ..., u_j)$ such that i is the smallest index for which $u_i = s$ and j is the largest index for which $u_j = t$. A *topological ordering of* G is an injective map $\mathcal{O} : N \to \mathbb{N}$ that assigns to every node a number such that, if two nodes are connected, the edge points from the lower to the higher numbered node. For $s, t \in N$ we will write $s < t$ to mean $\mathcal{O}(s) < \mathcal{O}(t)$ and the same for ">", when the topological ordering is implied. Given a topological ordering $\mathcal{O}$, the set of *predecessors* of a node s, denoted $\mathrm{pr}^{\mathcal{O}}(s)$, is the set of all nodes with a lower topological number, i.e. $\mathrm{pr}^{\mathcal{O}}(s) = \{t \in N : \mathcal{O}(t) < \mathcal{O}(s)\}$. For alternative DAGs G' or $\bar{G}$ we denote the above defined symbols with their respective accent, e.g. $\mathrm{ch}'(s), \bar{\mathrm{pa}}(s), \perp'_d, <'$, etc.

Probability on graphs

To every node $s \in N$ we associate a random variable X_s taking values in a finite state space X_s with typical elements x_s. For a subset $A \subset \mathsf{X}$ we let $\mathsf{X}_A = \times_{s \in A} \mathsf{X}_s$ and $\mathsf{X} = \mathsf{X}_N$. A typical element of X_A is denoted $x_A = (x_s)_{s \in A}$. The set of probability distributions over X is denoted $\mathcal{P}(\mathsf{X})$. If, for some distribution p, X_A is conditionally independent of X_B given X_C we will write $A \perp\!\!\!\perp B \mid C$.

For Y, Z finite sets, a *Markov kernel* is a map $k(\cdot \mid \cdot) : \mathsf{Y} \times \mathsf{Z} \to [0, 1]$ such that $\sum_{s \in \mathsf{Y}} k(s \mid t) = 1$, for all $t \in \mathsf{Z}$. A probability distribution $p \in \mathcal{P}(\mathsf{X})$ is said to *factorise* over G, if we can write the density as follows:

$$p(x) = \prod_{s \in N} k^s \left(x_s | x_{\mathrm{pa}(s)} \right). \tag{1}$$

where the k^s are Markov kernels. We denote the set of probability distributions over X that factorise over G by $\mathcal{P}^G$. A Markov kernel $k : \mathsf{X}_{N \setminus \mathrm{Roots}(G)} \times \mathsf{X}_{\mathrm{Roots}(G)} \to [0, 1]$ is said to *factorise* over G if it can be written as follows:

$$k(x_{N \setminus \mathrm{Roots}(G)} | x_{\mathrm{Roots}(G)}) = \prod_{s \in N \setminus \mathrm{Roots}(G)} k^s \left(x_s | x_{\mathrm{pa}(s)} \right) \tag{2}$$

We denote the set of such Markov kernels by $\mathcal{K}^G$.

Problem statement

Goal I Given a DAG $G = (N, E)$, find a DAG $G' = (N, E')$ such that $\mathrm{Roots}(G') = \mathrm{Leaves}(G)$ and for every $p \in \mathcal{P}^G$, the conditional distribution $p_{N \setminus \mathrm{Leaves}(G) | \mathrm{Leaves}(G)} \in \mathcal{K}^{G'}$.

It turns out (Proposition 4 in Appendix) that this goal is equivalent (up to edges between nodes in $\mathrm{Roots}(G')$) to the following goal:

Goal II Given a DAG $G = (N, E)$, find a DAG $G' = (N, E')$ such that there exists a topological ordering[1] of G' such that $\mathrm{Leaves}(G)$ are the eldest and $\mathcal{P}^{G'} \supset \mathcal{P}^G$.

[1]For necessity see Figure 4 in the Appendix.

In the remaining of the paper, we will focus on formulation II of the goal. Moreover we sometimes impose the following extra condition:

$$G' \supset G^*. \tag{3}$$

It can be argued that this is a natural condition since this enforces that the hierarchical structure of the generative model G is preserved when finding a suitable G'. Note that this condition also guarantees that there exists a topological ordering of G' such that Leaves(G) are eldest.

Preliminaries

Lemma 1. *If $G = (N, E)$ and $\bar{G} = (N, \bar{E})$ are such that $E \subset \bar{E}$ then $\mathcal{P}^G \subset \mathcal{P}^{\bar{G}}$.*

Proof. Since $\bar{\text{pa}}(s) \supset \text{pa}(s)$ for every node s, a distribution that can be written as $\prod_s k^s(x_s|x_{\bar{\text{pa}}(s)})$ can also be written as $\prod_s k^s(x_s|x_{\text{pa}(s)})$. ☐

Lemma 2. *Let A, B, S be subsets of N. We have, $A \perp\!\!\!\perp B \mid S$ for all $p \in \mathcal{P}^G$ if and only if S d-separates A and B in G.*

Proof. ($\Longleftarrow$) Corollary 5.11 and Proposition 5.13 in Cowell et al. (1999).
($\Longrightarrow$) Meek (1995) ☐

Lemma 3. *(Theorem 5.14 in Cowell et al. (1999)) Let G be a DAG with topological ordering $\mathcal{O}$. For a probability distribution p on X which has density with respect to a product measure, the following conditions are equivalent:*

(1) $p \in \mathcal{P}^G$;

(2) for all triples (A, B, S) of subsets of N such that $A \perp_d B \mid S$ we have $A \perp\!\!\!\perp B \mid S$ w.r.t. p;

(3) for all s we have $s \perp\!\!\!\perp \text{nd}(s) \mid \text{pa}(s)$ w.r.t. p;

(4) for all s we have $s \perp\!\!\!\perp \text{pr}^{\mathcal{O}}(s) \mid \text{pa}(s)$ w.r.t. p.

Corollary 1. *Let $\mathcal{O}, \mathcal{O}'$ be two topological orderings of G. If p satisfies property (4) of Lemma 3 w.r.t. $\mathcal{O}$, then the same is true for $\mathcal{O}'$.*

Proof. Note that $(1) - (3)$ of Lemma 3 are independent of the topological ordering. Therefore we have the following implications: for all s we have $s \perp\!\!\!\perp \text{pr}^{\mathcal{O}}(s) \mid \text{pa}(s)$ w.r.t. p $\Longrightarrow$ $p \in \mathcal{P}^G$ (with topological ordering $\mathcal{O}$) $\Longrightarrow$ $p \in \mathcal{P}^G$ (with topological ordering $\mathcal{O}'$) $\Longrightarrow$ for all s we have $s \perp\!\!\!\perp \text{pr}^{\mathcal{O}'}(s) \mid \text{pa}(s)$ w.r.t. p. ☐

In the rest of the paper, we fix a topological ordering for every DAG and in light of the the corollary, it doesn't matter which. Furthermore, we will omit the dependence on the topological ordering when talking about the set of predecessors.

Results

Conditions in terms of (d-)separation

Necessary and sufficient conditions for our goal can be deduced from the following theorem:

Theorem 1. *Let $G = (N, E), G' = (N, E')$ be DAGs. The following statements are equivalent:*

(1) $\mathcal{P}^{G'} \supset \mathcal{P}^{G}$

(2) For all sets A, B, S such that $A \perp'_d B \mid S$, we have $A \perp_d B \mid S$

(3) For all $s \in N$, we have $s \perp_d \mathrm{nd}'(s) \mid \mathrm{pa}'(s)$

(4) For all $s \in N$, we have $s \perp_d \mathrm{pr}'(s) \mid \mathrm{pa}'(s)$

Proof. (1) $\implies$ (2) (by contradiction) Suppose there exist A, B, S such that $A \perp'_d B \mid S$, but $A \not\perp_d B \mid S$. By Lemma 2 this implies there exists an $p \in \mathcal{P}^G$ for which $A \not\perp\!\!\!\perp B \mid S$. This violates (2) of Lemma 3 and therefore $p \notin \mathcal{P}^{G'}$.
(2) $\implies$ (1) Let $p \in \mathcal{P}^G$. We need to show $p \in \mathcal{P}^{G'}$. From (1) $\implies$ (2) in Lemma 3 we know that $A \perp_d B \mid S \implies A \perp\!\!\!\perp B \mid S$ for p. Combining this with the assumption $A \perp'_d B \mid S \implies A \perp_d B \mid S$ gives $A \perp'_d B \mid S \implies A \perp\!\!\!\perp B \mid S$. This means that p satisfies (2) Lemma 3 w.r.t. G' and therefore $p \in \mathcal{P}^{G'}$.
(1) $\iff$ (3) and (1) $\iff$ (4) can be shown in a similar way. □

Conditions in terms of perfectness

A sufficient condition for our goal can be deduced from the following theorem:

Theorem 2. *Let $G = (N, E), G' = (N, E')$ be two DAGs. If G' contains a subgraph $\bar{G}'$ such that $\bar{G}'$ is perfect and $\bar{G}'^{\sim} \supset G^{\mathrm{M}}$ then, $\mathcal{P}^{G'} \supset \mathcal{P}^{G}$.*

Proof. Let $p \in \mathcal{P}^G$. By Lemma 5.9 from Cowell et al. (1999) we know that p factorises undirectedly[2] over the undirected graph G^{M} and thus any undirected supergraph $H = (N, E^H)$ thereof. From Proposition 5.15 in Cowell et al. (1999) we know that p factorises (directedly) over any perfect directed graph $\bar{G}'$ such that $\bar{G}'^{\sim} = H$. Therefore when $\bar{G}'^{\sim} \supset G^{\mathrm{M}}$ we have $\mathcal{P}^{G'} \supset \mathcal{P}^{\bar{G}'} \supset \mathcal{P}^{G}$. □

From this theorem we can conclude that if we flip all the edges of G and then add edges until both G' is perfect and $G'^{\sim} \supset G^{\mathrm{M}}$, we obtain an inverse of G that satisfies our goal. The example in Figure 2 shows however that the condition that G' needs to contain a perfect subgraph $\bar{G}'$ such that $\bar{G}'^{\sim} \supset G^{\mathrm{M}}$ is not a necessary condition.

We do have the following necessary condition on the graph G' to satisfy our goal.

Theorem 3. *Let $\mathcal{P}^{G'} \supset \mathcal{P}^{G}$ and $G' \supset G^*$. For every s in N, the vertex induced subgraph $G'[\{s\} \cup \mathrm{des}(s)]$ contains a perfect subgraph $\bar{G}'_s$, such that $\bar{G}'^{\sim}_s \supset G^{\mathrm{M}}[\{s\} \cup \mathrm{des}(s)]$.*

[2] For a definition of this type of factorisation see Section 5.2 of Cowell et al. (1999)

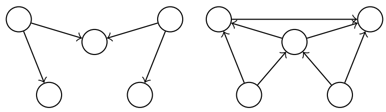


Figure 2: Pair of DAGs G, G' that satisfy Goal II but G' does not satisfy the condition in Theorem 2

Corollary 2. *If $\mathcal{P}^{G'} \supset \mathcal{P}^{G}$, $G' \supset G^*$, and $|\operatorname{Roots}(G)| = 1$, then G' contains a perfect subgraph $\bar{G}'$ such that $\bar{G}'^{\sim} \supset G^{\mathrm{M}}$.*

Note that this corollary implies that when $|\operatorname{Roots}(G)| = 1$ the conditions of Theorem 2 are both sufficient and necessary.
In order to prove Theorem 3, we first derive some preliminary results.

Definition 1. *Let G' be such that $\mathcal{P}^{G'} \supset \mathcal{P}^{G}$. We call an edge $e = (t, s)$ in G'* essential *if $\mathcal{P}^{G' \setminus \{e\}} \not\supset \mathcal{P}^{G}$.*

Proposition 1. *Let G, G' be such that $\mathcal{P}^{G'} \supset \mathcal{P}^{G}$ and $e = (s, t) \in E'$ such that s and t are connected in G^{M}. Then e is essential.*

Proof. First, suppose $(s, t) \in E$. Consider the distribution p for which $X_s = X_t$ and all other nodes (including s) are distributed independently according to Bernoulli(0.5). This is clearly in $\mathcal{P}^{G}$ but not in $\mathcal{P}^{G' \setminus \{e\}}$. The case $(t, s) \in E$ works similarly. Now suppose $(t, s) \notin E$. That means that t, s share at least one common child c. We consider the distribution where $X_c = X_s + X_t \mod 2$ and all other nodes (including s, t) are distributed independently according to Bernoulli(0.5). This is again clearly in $\mathcal{P}^{G}$ but not in $\mathcal{P}^{G' \setminus \{e\}}$. □

Proposition 2. *Let G, G' be such that $\mathcal{P}^{G'} \supset \mathcal{P}^{G}$ and $|\operatorname{Roots}(G)| = 1$. Then an edge $e = (t, s) \in E'$ is essential if and only if there exists a trail in G from t to s that is unblocked by $\operatorname{pa}'(s) \setminus \{t\}$.*

Proof. ($\Longleftarrow$) Suppose there exists a trail from t to s in G that is unblocked by $\operatorname{pa}'(s) \setminus \{t\}$. In $G' \setminus \{e\}$, t is no longer element of the parents of s but it remains to be a non-descendant. Therefore (3) of Theorem 1 is violated and we conclude $\mathcal{P}^{G' \setminus \{e\}} \not\supset \mathcal{P}^{G}$.
($\Longrightarrow$) If the edge e is essential there must be a $p \in \mathcal{P}^{G}$ for which there do not exist k's such that $p(x) = k^s(x_s | x_{\operatorname{pa}'(s) \setminus \{t\}}) \prod_{r \neq s} k^r(x_r | x_{\operatorname{pa}'(r)})$. This implies that $s \not\perp\!\!\!\perp t \mid (\operatorname{pa}'(s) \setminus \{t\})$ and therefore by the contrapositive of (1) $\Longrightarrow$ (2) of Theorem 3 there must be an unblocked trail. □

Proposition 3. *Let G, G' be such that $\mathcal{P}^{G'} \supset \mathcal{P}^{G}$, $G' \supset G^*$, and $|\operatorname{Roots}(G)| = 1$ and let $e_1 = (t_1, s), e_2 = (t_2, s)$ essential edges in G'. Then we have that t_1 and t_2 are connected in G'.*

Proof. Assume WLOG that $t_1 >' t_2$. By Proposition 2 we know that there exists trails $\gamma_1 : t_1 \to s, \gamma_2 : s \to t_2$ in G that are unblocked by $\operatorname{pa}'(s) \setminus \{t_1\}$ and $\operatorname{pa}'(s) \setminus \{t_2\}$ respectively. Note that we can assume WLOG that these trails do not contain loops.

Claim The path $\gamma \equiv \gamma_1 + \gamma_2$ is unblocked by $\mathrm{pa}'(t_1) \setminus \{t_2\}$.

In order to prove that γ is unblocked by $\mathrm{pa}'(t_1) \setminus \{t_2\}$, we need to show that for every edge u on the trail (1) if u is not a v-structure then $u \notin \mathrm{pa}'(t_1) \setminus \{t_2\}$, and (2) if u is a v-structure then u or one of its descendants in G is in $\mathrm{pa}'(t_1) \setminus \{t_2\}$. We start by considering the first case.

Case 1 Let u be an vertex in γ_i such that it is not a v-structure and unequal to s, t_i. Since γ_i is unblocked by $\mathrm{pa}'(s) \setminus \{t_i\}$, we have $u \notin \mathrm{pa}'(s)$. Furthermore, the trail γ_i restricted between u and s is unblocked by $\mathrm{pa}'(s)$, since t_i is not part of this trail by the no loop assumption. Therefore $s \not\perp_d u \mid \mathrm{pa}'(s)$. In order to satisfy condition (4) of Theorem 1 we need that $u \notin \mathrm{pr}'(s)$, i.e. $u >' s$. Therefore we have $u >' t_1$ and thus $u \notin \mathrm{pa}'(t_1)$.

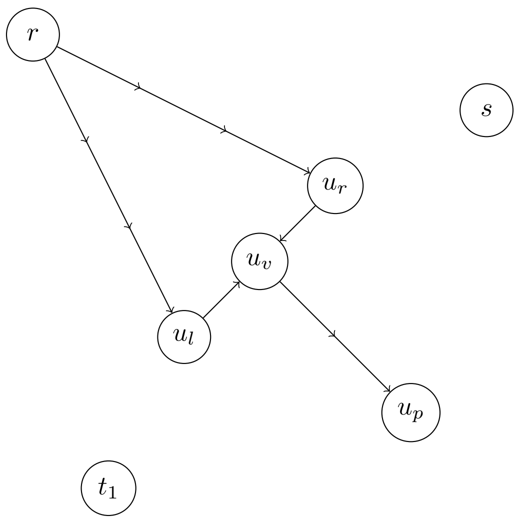


Figure 3: Relationship in G between different nodes introduced in the proof of Proposition 3

Case 2 Now let $u_l \to u_v \leftarrow u_r$ be the first v-structure in γ and j such that $u_v \in \gamma_j$. We need to show that u_v or one of its descendants in G is in $\mathrm{pa}'(t_1) \setminus \{t_2\}$ or modify the trail γ so that this v-structure is avoided.

(a) First assume $u_v \neq s$. Because γ_i is unblocked by $\mathrm{pa}'(s) \setminus \{t_i\}$, we know that u_v or a descendant (in G) of u_v is a parent of s in G' unequal to t_i. Let us denote the oldest node in G that is in $(\mathrm{pa}'(s) \setminus \{t_i\}) \cap (u_v \cup \mathrm{des}(u_v))$ by u_p (see Figure 3).

(i) $u_p = t_1$. In this case, we modify γ, such that $\gamma = (t_1 \leftarrow ... \leftarrow u_v) + \gamma_{u_v,t_2}$. Note that $(t_1 \leftarrow ... \leftarrow u_v)$ is unblocked by $\mathrm{pa}'(t_1) \setminus \{t_2\}$ since all vertices are younger than t_1 in G'

(ii) $u_p = t_2$. In this case, either $(t_2 \leftarrow ... \leftarrow u_v) \cap \mathrm{pa}'(t_1) \setminus \{t_2\}$ is non-empty, in which case we have what we need, or $t_2 \leftarrow ... \leftarrow u_v) \cap \mathrm{pa}'(t_1) \setminus \{t_2\} = \emptyset$ in which case we can modify γ, such that $\gamma = (t_2 \leftarrow ... \leftarrow u_v) + \gamma_{u_v,t_1}$. Since $(t_2 \leftarrow ... \leftarrow u_v)$ is now unblocked by $\mathrm{pa}'(t_1) \setminus \{t_2\}$, we have successfully avoided the v-structure.

(iii) $u_p <' t_1$ and $u_p \neq t_2$.

Lemma 4. *If $t_j \notin (u_p \to ... \to u_v)$, the trail in G from t_1 to u_p, given by:*

$$\bar{\gamma} := \gamma_{t_1,u_v} + (u_v \to ... \to u_p) \tag{4}$$

is unblocked by $\mathrm{pa}'(t_1) \setminus \{u_p\}$*. And therefore,* $t_1 \not\perp_d u_p \mid \mathrm{pa}'(t_1) \setminus \{u_p\}$.

Proof. Since u_v is the first v-structure on the trail, we know from Case 1 that γ_{t_1,u_v} is unblocked by $\mathrm{pa}'(t_1)$. If $u_v = u_p$ this proves the lemma. For $u_v \neq u_p$ we know that the trail $u_v \to ... \to u_p$ does not contain v-structures. Furthermore the nodes except u_p cannot be parents (in G') of s since u_p is the first descendant of u_v in G *and* parent of s in G'. Let $u_y \in (u_v \to ... \to u_p)$. Since $t_j \notin (u_v \to ... \to u_p)$ we know that the trail $(u_y \leftarrow ... \leftarrow u_v) + \gamma_{u_v,s}$ is unblocked by $pa'(s)$ (similar to Case 1). Therefore by the same reasoning as in Case 1, $u_y >' s$ and can therefore not be a parent (in G') of t_1. Therefore also in this case $\bar{\gamma}$ is unblocked by $\mathrm{pa}'(t_1) \setminus \{u_p\}$. □

From Lemma 4 and (4) in Theorem 1 we have that if $t_j \notin (u_v \to ... \to u_p)$ then $u_p \in \mathrm{pa}'(t_1)$. In case $t_j \in (u_v \to ... \to u_p)$ we can modify γ_j such that $\gamma_j = (t_j \to ... \to u_v) + \gamma_{u_v,s}$. This trail is unblocked by $\mathrm{pa}'(s) \setminus \{t_j\}$ and therefore the v-structure is successfully avoided.

(iv) $u_p >' t_1$. In this case also the nodes in $(r \to ... \to u_l)$ and $(r \to ... \to u_l)$ are younger than t_1 in G' and therefore not in $\mathrm{pa}'(t_1)$. Now we can modify the path as follows so that the v-structure is bypassed: $\gamma_{t_1,u_l} + (u_l, ..., r) + (r, ..., u_r) + \gamma_{u_r,t_2}$.

(b) When s itself is a v-structure in γ then we can introduce a similar bypass as in (iv) using that $s >' t_1$.

By iteratively applying the modifications when necessary we obtain a trail in G between t_1 and t_2 that is unblocked by $\mathrm{pa}'(t_1) \setminus \{t_2\}$ and therefore the edge (t_2, t_1) is essential in G'. □

Lemma 5. *Let $A \subset N$. If $G = (N, E)$ and $\bar{G} = (N, \bar{E})$ are such that $\mathcal{P}^G \subset \mathcal{P}^{\bar{G}}$, then the same holds for the vertex induces subgraph of both graphs: $\mathcal{P}^{G[A]} \subset \mathcal{P}^{\bar{G}[A]}$*

Proof. One can easily check that the condition (3) in Theorem 1 remains satisfied when taking vertex induced subgraphs. □

Proof of Theorem 3. Let us first assume $|\mathrm{Roots}(G)| = 1$ and consider the subgraph $\bar{G}' \subset G'$ with only essential edges. From Proposition 1 we know that $\bar{G}' \supset G^{\mathrm{M}}$. From Proposition 3 we know that $\bar{G}'$ is perfect. Therefore by Theorem 2 we know $\mathcal{P}^{\bar{G}'} \supset \mathcal{P}^G$ and thus by Lemma 1 $\mathcal{P}^{G'} \supset \mathcal{P}^G$.
Now we consider a general DAG G with $|\mathrm{Roots}(G)| \geq 1$. Note that by Lemma 5 for a fixed $s \in N$, $\mathcal{P}^{G'} \supset \mathcal{P}^G$ implies $\mathcal{P}^{G'[\{s\} \cup \mathrm{des}(s)]} \supset \mathcal{P}^{G[\{s\} \cup \mathrm{des}(s)]}$. Since s is the unique root for $G[\{s\} \cup \mathrm{des}(s)]$, we know from the above case that this implies that $G'[\{s\} \cup \mathrm{des}(s)]$ contains a perfect subgraph $\bar{G}'_s$, such that $\bar{G}'_s \supset G^{\mathrm{M}}[\{s\} \cup \mathrm{des}(s)]$ is necessary. Since s was chosen arbitrarily this proves the result. □

In practice, the inverse G' is often obtained by simply inverting the edges in G. In this case we have the following necessary and sufficient condition to satisfy our goal.

Theorem 4. *If* $G' = G^*$, *then:* $\mathcal{P}^{G'} \supset \mathcal{P}^{G} \iff \mathrm{pa}(s), \mathrm{ch}(s)$ *are complete for all* $s \in N$.

Proof. ($\Longleftarrow$) If $\mathrm{pa}(s), \mathrm{ch}(s)$ are complete for all $s \in N$ and $G' = G^*$ this implies that $G'^{\sim} \supset G^{\mathrm{M}}$ and G' is perfect. The result now follows from Theorem 2. ($\Longrightarrow$) We will show the contrapositive. Assume first that there exists an $s \in N$ such that $t_1, t_2 \in \mathrm{pa}(s)$ are not connected. Now consider the distribution $p \in \mathcal{P}^G$ for $X_s = X_{t_1} + X_{t_2} \mod 2$ and all other nodes are Bernoulli(0.5). It is easy to see that $p \notin \mathcal{P}^{G'}$. Now assume that there exists an $s \in N$ such that $u_1, u_2 \in \mathrm{ch}(s)$ are not connected. Now consider the distribution $p \in \mathcal{P}^G$ such that x_{u_1} and X_{u_2} are equal to X_s and all other nodes (including s itself) are Bernoulli(0.5). It is again easy to see that $p \notin \mathcal{P}^{G'}$. □

Conclusion

In this paper, we derived necessary and sufficient conditions for the recognition network to be able to model the exact posterior distribution of a generative Bayesian network. In case that the generative network has a single node without the parents, the necessary and sufficient conditions coincide. However, for multiple nodes without parents there is still a gap in both conditions. The authors would like to pursue this question further to find a single necessary and sufficient condition also for the general case.
As mentioned in the introduction, the results in this paper do not cover the case in which the distributions of individual nodes are further restricted, for example by fixing them to be Gaussian, with mean and variance being a function of the values of the parent nodes. A further direction of study would be to investigate the relationship between the architecture of the recognition network and its ability to model the posterior, in the case of the restricted set of probability distributions.

References

R. G. Cowell, P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer-Verlag New York, 1999.

P. Dayan, G. E. Hinton, R. M. Neal, and R. S. Zemel. The helmholtz machine. *Neural computation*, 7(5):889–904, 1995.

I. Flesch and P. J. Lucas. Markov equivalence in bayesian networks. In *Advances in probabilistic graphical models*, pages 3–38. Springer, 2007.

D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

S. L. Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

C. Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 403–410, 1995.

T. Verma and J. Pearl. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, pages 255–270, 1990.

M. J. Wainwright, M. I. Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

Appendix

Multiple possible topological orderings where only one satisfies the goal

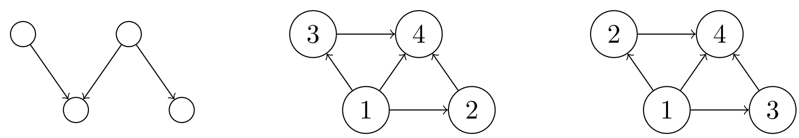


Figure 4: Pair of DAGs G, G' for which there is freedom to choose the topological ordering

Equivalence of two goals

Proposition 4. *Let* $G = (N, E)$ *be a DAG,* $\tilde{G} = (N, E \cup \tilde{E}), \tilde{E} = \{(s,t) : s,t \in \mathrm{Roots}(G), s < t\}$*, and* $\mathcal{S} \subset \mathcal{P}(\mathsf{X})$.

$$\forall p \in \mathcal{S}, p_{N\setminus \mathrm{Roots}(G)|\,\mathrm{Roots}(G)} \in \mathcal{K}^G \iff \mathcal{P}^{\tilde{G}} \supset \mathcal{S}. \tag{5}$$

Proof. ($\implies$) Let $p \in \mathcal{S}$ and suppose $p_{N\setminus \mathrm{Roots}(G)|\,\mathrm{Roots}(G)} \in \mathcal{K}^G$. We need to show $p \in \mathcal{P}^{\tilde{G}}$. We can write p as follows:

$$p(x) = p(x_{N\setminus \mathrm{Roots}(G)}|x_{\mathrm{Roots}(G)})p(x_{\mathrm{Roots}(G)}) \tag{6}$$

From the fact that $p_{N\setminus \mathrm{Roots}(G)|\,\mathrm{Roots}(G)} \in \mathcal{K}^G$ and all the nodes in $\mathrm{Roots}(G)$ are connected in $\tilde{G}$, this can be written as follows:

$$p(x) = \prod_{s \in N\setminus \mathrm{Roots}(G)} k^s(x_s|x_{\mathrm{pa}(s)}) \prod_{s \in \mathrm{Roots}(G)} k^s(x_s|x_{\tilde{\mathrm{pa}}(s)}) \tag{7}$$

$$= \prod_{s \in N} k^s(x_s|x_{\tilde{\mathrm{pa}}(s)}), \tag{8}$$

and therefore $p \in \mathcal{P}^{\tilde{G}}$.

($\Longleftarrow$) Now let $p \in \mathcal{S}$ again and suppose $p \in \mathcal{P}^{\tilde{G}}$. We can write

$$p(x) = \prod_{s \in N} k^s(x_s | x_{\tilde{\text{pa}}(s)}) \tag{9}$$

$$= \prod_{s \in N \setminus \text{Roots}(G)} k^s(x_s | x_{\tilde{\text{pa}}(s)}) \prod_{s \in \text{Roots}(G)} k^s(x_s | x_{\tilde{\text{pa}}(s)}) \tag{10}$$

$$= \prod_{s \in N \setminus \text{Roots}(G)} k^s(x_s | x_{\text{pa}(s)}) \prod_{s \in \text{Roots}(G)} k^s(x_s | x_{\tilde{\text{pa}}(s)}) \tag{11}$$

where we can switch from $\tilde{\text{pa}}$ to pa in the third equality because there are only edges added between nodes in Roots(G) to obtain $\tilde{G}$. It can be shown that $\prod_{s \in \text{Roots}(G)} k^s(x_s | x_{\tilde{\text{pa}}(s)}) = p(x_{\text{Roots}(G)})$. Dividing by $p(x_{\text{Roots}(G)})$ on both sides gives:

$$p(x_{N \setminus \text{Roots}(G)} | x_{\text{Roots}(G)}) = \prod_{s \in N \setminus \text{Roots}(G)} k^s(x_s | x_{\text{pa}(s)}) \tag{12}$$

and therefore $p_{N \setminus \text{Roots}(G) | \text{Roots}(G)} \in \mathcal{K}^G$. $\square$

Concurrent Business and Distribution Strategy Planning Using Bayesian Networks

Theodor Petřík[1] and Martin Plajner[2]

[1]Faculty of Social Sciences, Institute of Economic Studies, Charles University,
Opletalova 38, Prague 1, 110 00, Czechia
[2]Institute of Information Theory and Automation, Czech Academy of Sciences,
Pod vodárenskou věží 4, Prague 8, 182 08, Czechia

Abstract

Business and distribution strategy planning are usually carried out in a sequence. A company first devises a business plan and then a distribution strategy able to accommodate it. The separation in planning can lead to a sub-optimal choice. We propose a method how to concurrently plan both strategies, using a Bayesian network. We present three modifications of our concurrent optimization model which are based on different optimization objectives - distribution strategy costs minimization, revenue maximization and profit maximization. The derivation of all model modifications and the collection process of the required inputs are described in detail.

The presented model is tested on a business case of the company Pilsner Urquell. Using the company historical data from 01/2017 – 12/2017, we design the cost optimum distribution strategy in the Czech market for years 2018 - 2020. Our results are then compared with the real company development over the same period. With our model we show that the company could have selected a more cost-effective distribution strategy in 2017.

1 Introduction

This paper introduces Bayesian Networks (BNs) (Jensen and Nielsen, 2007) as an effective tool for concurrent business and distribution strategy planning. Planning of these two company processes is usually performed in sequence when the business plan is created first. However, the two processes are inherently connected and omitting their dependence can lead to a sub-optimal strategy selection. Therefore, it is necessary to conduct planning in both domains concurrently.

Every company engages in a form of business planning. The methods most frequently debated in the literature are Trend-Impact Analysis (Gordon and Stover, 1976) , Cross-Impact Analysis (Gordon and Stover, 2003) and Intuitive Logics (R. Huss and J. Honton, 1987). These approaches usually rely on an expert opinion more than historical data.

The objective of a distribution network design is to plan the most cost efficient manner of product movement through the whole supply chain (Ambrosino and Grazia Scutellà, 2005). To stress the importance, Ballou (2001) estimated that through an efficient distribution network and effective facility management, the operation costs can be decreased by up to 15%. Mangiaracina et al. (2015) composed a highly comprehensive review of the distribution network optimization methods in the contemporary literature. However, there is no methodology available for concurrent planning together with a business plan.

In our work, we use BNs as the underlining probabilistic model. BNs are already known in many other tasks in supply chain, for example, supplier selection (Hosseini and Barker, 2016) or more general, supplier resilience and risk management (Sharma et al., 2022).

The article is structured as follows. First, we establish our notation in Section 2. Next, we develop the concurrent optimization model, the main contribution of this article, in Section 3. The model is then implemented in Section 4 on a business case of Pilsner Urquell where we use it to plan an optimum long-term distribution strategy. The last Section 5 provides an overview and a potential for further research.

2 Notation and preliminary steps

In this article, we use the BNs as an effective tool to concurrently plan business and distribution strategies. Importantly, our method does not replace the existing approaches for business nor distribution planning processes. It represents an extension allowing to effectively combine and evaluate the information from both processes. This section describes the process of inputs collection for our model and establishes the variables that we use.

The preliminary step is to gather the inputs from the business and distribution planning processes. First, we describe the step for the business planning process. Although the usability of our approach is not conditioned by use of any specific planning method, outputs from the process must be transferable to the to the BN. An example of our BN structure is shown in Figure 1. We work with a time outlook for n consecutive periods.

- Variable A^i, $i \in 1 \dots n$, is the modeled company in the time period i and its states $a^i_j, j \in 1 \dots m^i$ are possible states the company can have in that time period. $\mathbf{A} = \{A^1, \dots, A^n\}$ is the set of all company nodes at all time periods.
- Variables $Y^i_q, i \in \{1, \dots, n\}, q \in \{1, \dots, s^i\}$ represent other events influencing the company. The subscript k specifies that there are s^i parent nodes Y for a period A^i.

Now we can proceed to the collection of inputs from the distribution network planning process. The company designs a number d of feasible distribution networks $\mathbf{Z}$ which

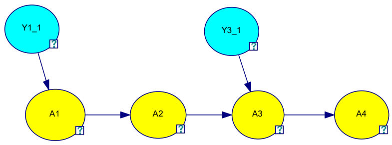


Figure 1: Example structure of our BN model

could accommodate the needs of the company $\{A^1, \ldots, A^n\}$. Symbol $Z_f^i, i \in \{1, \ldots, n\}$, $f \in \{1, \ldots, d\}$ then refers to a strategy Z_f implemented during a specific period i.

Next, we estimate several key performance indicators (KPIs) which we implement in our model. KPIs are metrics that companies track to measure their performance. Specifically, we implement revenue r, distribution network operating costs c and profit p.

- Revenue is the total amount of income generated by the sales of goods and services that a company provides. In our model, $r_{j,f}^i, i \in \{1, \ldots, n\}, f \in \{1, \ldots, d\}$, $j \in \{1, \ldots, m^i\}$ stands for revenue in a state a_j^i while operating a distribution network Z_f. The tool selected to estimate $r_{j,f}^i$ can be chosen freely but it must be capable of doing so for every Z_f at every state a_j^i included in the BN model.
- Distribution network operating costs for a company are costs directly related to the network operations (fuel, worker wages, ...). In our model, $c_{j,f}^i, i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m^i\}, f \in \{1, \ldots, d\}$ stands for distribution network operating costs in a state a_j^i while operating a distribution network Z_f. The tool selected to estimate $c_{j,f}^i$ can be chosen freely but it must be capable of doing so for every Z_f at every state a_j^i included in the BN model.
- We define the profit p as the difference between r and c, hence

$$p_{j,f}^i = r_{j,f}^i - c_{j,f}^i$$

.

The dependency of both r and c on a_j^i and Z_f is a logical consequence of our empirical experience and fundamental belief, that the business and distribution planning processes are dependent and should be planned concurrently.

Finally, we need to estimate the transition costs $t_{e,f}^i, i \in \{2, \ldots, n\}, e, f \in \{1, \ldots, d\}$. Transition costs $t_{e,f}^i$ is an additional expense a company must make between periods $i-1$ and i, when changing from a Z_e to Z_f. These costs can be, for example, associated with moving from an existing facility to a new facility. An estimate must be made for every pair due to the possibility of the following situation: $t_{e,f}^i \neq t_{f,e}^i, e, f \in \{1, \ldots, d\}$.

Conditional probabilities of A^i are all obtained from the preliminary planning processes. They are as follows.

$$P(A^i | Pa(A^i)), i \in \{1, \ldots, n\}$$

2.1 Estimation of costs for every scenario using DW

To estimate the designed distribution networks across all business scenarios, we use a simulation software Distribution wizard (DW), developed by company Logio. This specialized simulation software implements an open-source engine JSprit[1] to model complex business scenarios. Effectiveness of the engine has been successfully demonstrated on a range of research activities to solve a variety of problems (Mahmoud et al., 2022). In the simulation process, DW incorporates a set of parameters allowing to realistically model a wide range of networks and to provide answers to many business questions.

JSprit uses a so called Ruin and Recreate (RAR) metaheuristic introduced by Schrimpf et al. (2000). They tested their new approach on the existing library of Vehicle Routing Problems (Dantzig and Ramser, 1959) (VRPs) and recorded overwhelmingly better results than what any other contemporary method could achieve.

Generally, the RAR principle contains three steps. First, it is necessary to define an admissible solution to the problem at hand which will obey all the predefined constraints. An admissible solution is a collection of routes (a sequence of jobs) to the given VRP. The *Ruin and Recreate* is a second step which attempts finding a better solution. The *Ruin* part selects a segment of the solution devised in the previous step and removes it from the solution. Subsequently, the removed part is recreated by the *Recreate* part as good as possible by again finding an admissible solution which obeys all the constraints. Lastly, the algorithm compares the new solution with the existing one and decides whether to preserve the previous, or keep the new solution.

3 Concurrent optimization model

In this section we present our model. First, we define a sequence of indices $\mathbf{k}_l = \{k_l^1, \ldots, k_l^n\}, k_l^i \in \{1, \ldots, d\}$ which creates a sequence of distribution networks in time periods $\mathbf{X}_l = \{Z_{k_l^1}, \ldots, Z_{k_l^n}\}$, where each l marks a single permutation of indices[2]. The goal of our model is to find an optimum sequence of $\mathbf{X}_l$ for the whole outlook $1 \ldots n$. We propose three modifications of our model, based on three different optimization objectives.

- **Distribution strategy costs minimization**

  The customer delivery costs and the warehousing costs are minimized while keeping the defined service level. Therefore, the objective of this problem is to find a sequence of $\mathbf{X}_l$, such, that in combination with the associated $t^i_{e,f}$, the sequence yields the lowest expected total operation costs.

- **Revenue maximization**

  Revenue maximization is a common objective that many firms pursue as a priority, for example, to increase their market share. The objective of this problem is to find

[1] https://github.com/graphhopper/jsprit

[2] There are n periods and d possible distribution network each period. Therefore, there is a total of d^n permutations.

a sequence of $\mathbf{X}_l$, such, that in combination with the associated $t^i_{e,f}$, the sequence yields the highest expected total revenue.

- **Profit maximization**

  The objective of this problem is to find a sequence of $\mathbf{X}_l$, such, that in combination with the associated $t^i_{e,f}$, the sequence yields the highest expected total profit.

3.1 Distribution strategy costs minimization

The first step is to obtain $\forall i \in \{1 \ldots n\}$ and $\forall f \in \{1 \ldots d\}$ the expected operation costs $\boldsymbol{\gamma^i_f}$. In total, there are $n * d$ estimates because we are estimating γ of every $Z_f, f \in \{1, \ldots, d\}$ at every $A^i, i \in \{1, \ldots, n\}$.

$$E[\gamma^i_f] = \sum_{j=1}^{m^i} c^i_{j,f} P(a^i_j | Pa(A^i)) \tag{1}$$

Using $E[\gamma^i_f]$ and $t^i_{e,f}$ we can now define the optimization problem.

$$\arg\min_l \{\sum_{i=1}^{n} E[\gamma^i_{k^i_l}] + \sum_{i=2}^{n} t^i_{k^{(i-1)}_l, k^i_l}\} \tag{2}$$

The resulting l is such a sequence of distribution networks which yields the minimum costs.

3.2 Revenue maximization

The first step is to obtain $\forall i \in \{1, \ldots, n\}$ and $\forall f \in \{1, \ldots, d\}$ the expected revenue $\boldsymbol{\rho^i_f}$. In total, there are $n * d$ estimates because we are estimating ρ of every $Z_f, f \in \{1, \ldots, d\}$ at every $A^i, i \in \{1, \ldots, n\}$.

$$E[\rho^i_f] = \sum_{j=1}^{m^i} r^i_{j,f} P(a^i_j | Pa(A^i)) \tag{3}$$

Using $E[\rho^i_f]$ and $t^i_{e,f}$ we can now define the optimization problem.

$$\arg\max_l \{\sum_{i=1}^{n} E[\rho^i_{k^i_l}] - \sum_{i=2}^{n} t^i_{k^{(i-1)}_l, k^i_l}\} \tag{4}$$

The resulting l is such a sequence of distribution networks which yields the maximum revenue.

3.3 Profit maximization

First step is to obtain $\forall i \in \{1, \ldots, n\}$ and $\forall f \in \{1, \ldots, d\}$ the expected total profit $\boldsymbol{\pi^i_f}$. In total, there are again $n * d$ estimates.

$$E[\pi^i_f] = \sum_{j=1}^{m^i} p^i_{j,f} P(a^i_j | Pa(A^i)) \tag{5}$$

Using $E[\pi^i_f]$ and $t^i_{e,f}$ we can now define the optimization problem.

$$\arg\max_l \{ \sum_{i=1}^{n} E[\pi^i_{k^i_l}] - \sum_{i=2}^{n} t^i_{k^{(i-1)}_l, k^i_l} \} \tag{6}$$

4 Case study

4.1 Introduction

We tested the method proposed in previous sections using data and business case of the company Pilsner Urquell. Pilsner Urquell Brewery (PU) is the largest brewery in Czechia, headquartered in Pilsen. PU has three production plants where beer and other beverages are produced. Its customers are large supermarket chains, smaller convenient stores, restaurants, and pubs. To help accommodate this vast network of clients, PU runs a network of fourteen strategically located depots in the Czech region.

PU runs its logistic at the high end of the domain standard and it achieves remarkable efficiency and results. This is possible because of their proper planning and long-term evaluations. The business task described below is on of the cases where PU wanted to analyze the situation on the market in advance and to be ready for the change when it arrives. This is necessary as all changes in logistics operations take a long time to implement. The goal is to set the optimum long-term distribution strategy in the Czech market for years 2018 - 2020.

The case study is structured as follows. First we provide the key facts regarding the company's operations and the business outlook. Next, we construct the BN model based on the business outlook and the historical data and simultaneously propose several feasible distribution strategies. Further, we evaluate each distribution strategy for each business scenario in terms of operation costs[3], using a specialized software created by the company Logio called Distribution wizard (DW). Consequently, using the outputs from DW and the estimated transition costs among strategies, we select the optimal distribution plan for the company for years 2018 - 2020.

Operations description

The company's distribution network can be divided into three transportation channels:

[3]Prices are always listed in units, corresponding to CZK*coefficient due to a confidentiality policy of PU. Therefore, all conclusions are expressed in relative terms which remain accurate.

- **Primary -** The primary channel is concerned with goods redistribution among the PU's facilities, mainly from the production plants to the depots. These shipments are almost always large amounts carried by trailers with $38t$ capacity.
- **Direct -** Through the direct channel, the product is delivered from the production plants to the distribution centers (DCs) operated by some large supermarket chains. These are always wholesale shipments carried by trailer with $24t$ capacity. This channel is the most cost effective because the product is transported in bulk to the customer using the most direct way.
- **Secondary -** Through the secondary channel, the product is delivered to all customers except those already delivered by the direct channel. These shipments are usually distributed using smaller trucks with $9.7t$ capacity.

4.2 Concurrent optimization

Business outlook

PU delivers its products to multiple supermarkets chains. Some supermarket chains already are delivered by the direct channel in 2017. Some of the large customers prefer the direct channel as it allows them to consolidate goods and to better manage the supply of their stores. Three additional chains are signalling a possibility of a request to change to this model as well. This change would result in a transition of a portion of deliveries from the secondary and primary channel to the direct channel. Also, PU does not expect a significant sales growth in the domestic market hence we assume the sales stay constant over the whole outlook.

Data

We obtained data from PU related to their distribution network operations for the full calendar year 2017. The VRP problem is very complex and the simulations require a lot of time to complete. Therefore, we select two calendar months, January and June on which the method is demonstrated. The two months are a representative sample. As can be seen in Figure 2, January is the slowest month of the year and June is when the summer peak occurs.

Bayesian network construction

Using the business outlook and the obtained data, we could proceed to the BN construction. Figure 3 shows the structure of our BN model. The model is built around the expectation that up to three customers will change to the direct way of delivery. There are three nodes $2018, 2019$ and 2020 representing the company over the three year outlook horizon, each having eight states. The states stand for every possible scenario when none or maximum of all three customers (A, B, C) shift their preference toward the direct delivery.

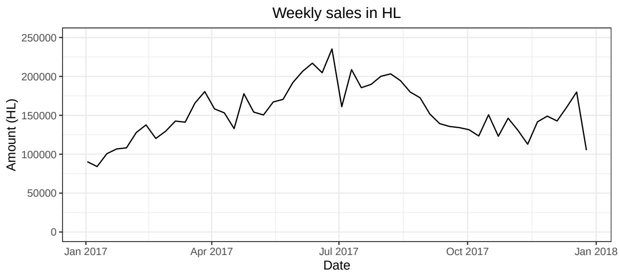


Figure 2: Monthly sales in HL

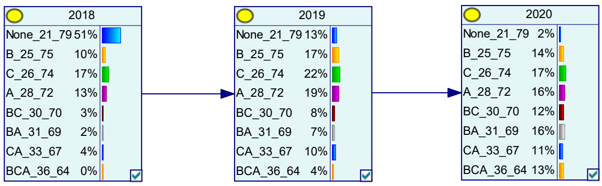


Figure 3: Bayesian network model

The name of every state also contains the respective shares of the direct and secondary distribution channels. For example, the state BA_31_69 represents the situation when customers B and A change to the direct distribution channel, resulting in a share of direct distribution of 31% and the share of secondary 69% [4]. The probability distribution on each node in the BN is based on the company expectations.

Possible distribution strategies

With respect to the business outlook, the company is considering to close the operations in one or two depots. Due to their location and throughput, depots in Teplice and Jihlava were the most fitting ones to be closed. There are therefore four distinct strategies in total which the company could adopt:

- No depot closes
- Depot in Teplice closes - based on the distance, the customers get split between the nearby depots in Karlovy Vary and Mnichovo Hradiště
- Depot in Jihlava closes - the customers get split among the nearby depots in Hradec Kralove, Praha, Ceske Budejovice and Brno
- Both depots close

[4] Primary is exempted because it is a redistribution among the PU facilities, never customer delivery

	No depot		Teplice		Jihlava		Both	
BS	**Jan**	**June**	**Jan**	**June**	**Jan**	**June**	**Jan**	**June**
Base	6,183	10,224	6,255	10,399	6,588	10,815	6,648	10,957
B	5,854	9,963	6,026	10,163	6,452	10,598	6,632	10,815
C	5,874	9,942	6,046	10,170	6,448	10,685	6,628	10,873
A	5,769	10,007	5,950	10,195	6,416	10,710	6,576	10,902
CB	5,789	9,713	6,002	9,924	6,516	10,486	6,721	10,710
AB	5,609	9,583	5,813	9,855	6,275	10,406	6,524	10,678
CA	5,737	9,688	5,902	10,029	6,400	10,569	6,592	10,815
CAB	5,420	9,282	5,665	9,557	6,187	10,253	6,432	10,547

Table 1: Operating costs of different network configurations (thous. units) as estimated by DW

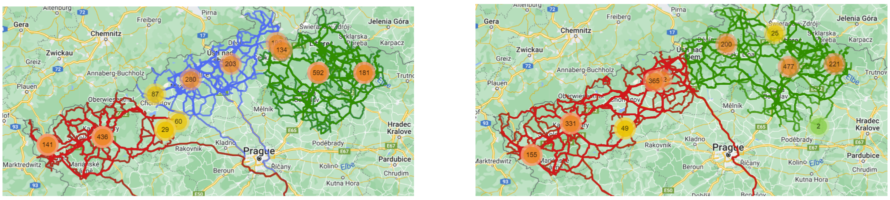
Closing a depot eliminates the fix and operation costs required to run it. However, closing a depot also results in an increased distance to the affected customers which in turn increases the distribution network operation costs in other depots.

4.3 Optimum strategy selection

We have constructed the BN with our future expectation (Figure 3) and estimated the operating costs of four possible distribution strategies (Table 1) at each state in the BN. Before we can proceed to find the optimum sequence of distribution strategies, we need to establish the transition costs and also adjust the operations costs estimates in Table 1 for the savings achieved by closing the depots.

The transition costs are extra, one time expenses which PU would have to make to change the network set up. We estimate the one time cost of closing the depot in Teplice to be 350,000 units and 550,000 units in Jihlava. We assign a large penalty to the cases when a closed depot would be reopened again because from the business perspective, it is an unrealistic development. Furthermore, PU calculated the Teplice depot operation costs for January and June to be 509,000 units and 671,000 units respectively. For the depot in Jihlava, the costs are higher at 911,000 units and 1,036,000 units. Having obtained all necessary inputs, we can apply our distribution strategy cost minimization model as described in Section 3.1. Using Formulas 1 and 2, we obtain the optimum sequence of distribution strategies for all three years.

Table 2 shows five transition paths with the lowest operation costs. For both months, the estimated optimum transition paths is identical, *Both-Both-Both*. The interpretation of this distribution plan is that PU should immediately close both depots in Teplice and in Jihlava and keep them closed for the whole outlook. Figures 4 and 5 show the situation before and after closing the depots. Although there are only minor costs differences among the top transition paths, the results clearly show that closing one or both depots would be beneficial for the company.



(a) Base setup with Teplice depot operational

(b) Alternative setup when depot in Teplice is closed

Figure 4: Closing the the Teplice depot

(a) Base setup with Jihlava depot operational

(b) Alternative setup when depot in Jihlava is closed

Figure 5: Closing the Jihlava depot

4.4 Comparison to the company data

We expected the company to have 0% growth over the outlook. In 2018 and 2019, PU sales grew yearly for less than 1%. Although, the growth rate reached 4% in 2020, our expectation was overall reasonably accurate.

For the base scenario in our model, we estimated the operation costs to be 6,183,000 units in January and 10,224,000 units in June. PU's real costs were, on average, higher at 7,821,000 units in January and 11,783,000 units in June. The difference can be explained as a potential between the near optimal state and the reality. In the near optimum, the trucks are always fully loaded and always choose the shortest path to complete the delivery. The relative differences of 26% in January, which is off-season, and 15% in June, during the summer peak, suggest that PU operates comparably a very efficient distribution network. The potential can be mainly found in the smaller depots with smaller customer base, for which it is much more demanding to achieve similar efficiency rates to the large depots, especially off-season.

Over 2018 - 2020, PU preserved all of their depots operational. The optimum distribution strategy selected by our model is closing both depots in Teplice and in Jihlava. As shown by our model, this strategy better corresponds with the company's persistent expectations regarding the changing preferences of the large customers. Closing these depot could have resulted in net savings of 5.5% over 2018-2020.

January		June	
Strategy	**Costs**	**Strategy**	**Costs**
Both-Both-Both	16,466	Both-Both-Both	28,288
Jihlava-Both-Both	16,856	Teplice-Both-Both	28,690
Teplice-Both-Both	16,871	Teplice-Teplice-Teplice	28,735
Teplice-Teplice-Teplice	16,876	Jihlava-Both-Both	28,786
Jihlava-Jihlava-Jihlava	17,157	Teplice-Teplice-Both	29,015

Table 2: The top five optimum transition paths for 2018-2019-2020 by January and June, based on the projected distribution strategy operating costs (thous. units)

5 Conclusion

Based on our empirical experience in the field, we believe that the separation in planning of business and distribution strategies can lead to a sub-optimal choice. In this article, we presented a new method for concurrent business and distribution strategy planning using a Bayesian network. The method was described and applied on a business case of the company Pilsner Urquell. Using our method, we selected the most cost effective distribution strategies for the company (Table 2). The company could have decrease its expected distribution network operation costs, had it followed the strategy selected by our model.

The area of concurrent business and distribution planning is yet unexplored and the purpose of this article was to introduce the field. In our future work we want to conduct a more comprehensive research in this area. Specifically, we want to further improve the precision of our model by including continuous variables in the BN. Also, we aim to generalize our method to ensure its applicability to other concurrent optimization problems.

Acknowledgements

This work was supported by the Czech Science Foundation (Project No. 19-04579S).
This article uses a prototype of distribution strategy modeling tool which is being developed by the company Logio with the support of Ministry of Industry
MPO 610128/21/61200/1051.

References

D. Ambrosino and M. Grazia Scutellà. Distribution network design: New problems and related models. *European Journal of Operational Research*, 165(3):610–624, Sept. 2005. ISSN 03772217.

R. H. Ballou. Unresolved Issues in Supply Chain Network Design. *Information Systems Frontiers*, pages 417–426, 2001.

G. B. Dantzig and J. H. Ramser. The truck dispatching problem. *Management science*, 6(1):80–91, 1959.

T. J. Gordon and J. Stover. Using perceptions and data about the future to improve the simulation of complex systems. *Technological Forecasting and Social Change*, 9(1-2): 191–211, 1976.

T. J. Gordon and J. G. Stover. Cross-Impact Analysis. *Futures research methodology*, (2), 2003.

S. Hosseini and K. Barker. A Bayesian network model for resilience-based supplier selection. *International Journal of Production Economics*, 180:68–87, Oct. 2016. ISSN 0925-5273.

F. V. Jensen and T. D. Nielsen. *Bayesian networks and decision graphs.* Springer, 2 edition, 2007.

A. Mahmoud, T. Chouaki, S. Hörl, and J. Puchinger. *Extending JSprit to solve electric vehicle routing problems with recharging.* Feb. 2022.

R. Mangiaracina, G. Song, and A. Perego. Distribution network design: a literature review and a research agenda. *International Journal of Physical Distribution & Logistics Management*, 45(5):506–531, June 2015. ISSN 0960-0035.

W. R. Huss and E. J. Honton. Scenario planning - What style should you use? *Long range planning*, 20(4):21–29, 1987. ISSN 0024-6301. doi: 10.1016/0024-6301(87)90152-X.

G. Schrimpf, J. Schneider, and Y. H. Stamm-wilbr. Record breaking optimization results using the ruin and recreate principle. *Journal of Computational Physics*, pages 139–171, 2000.

S. K. Sharma, S. Routroy, and U. Chanda. Supply-side risk modelling using Bayesian network approach. *Industrial Management and Data Systems*, 112:24, 2022.

Continuity and Additivity Properties of Information Decompositions

Johannes Rauh[1], Pradeep Kr. Banerjee[1], Eckehard Olbrich[1], Guido Montúfar[1,2], and Jürgen Jost[1]

[1]Max Planck Institute for Mathematics in the Sciences
{jrauh,pradeep,olbrich,jjost}@mis.mpg.de
[2]University of California, Los Angeles
montufar@math.ucla.edu

Abstract

Information decompositions quantify how the Shannon information about a given random variable is distributed among several other random variables. Various requirements have been proposed that such a decomposition should satisfy, leading to different candidate solutions. Curiously, however, only two of the original requirements that determined the Shannon information have been considered, namely monotonicity and normalization. Two other important properties, continuity and additivity, have not been considered. In this contribution, we check which of the decompositions satisfy these two properties. While most of them satisfy continuity, only one of them satisfies additivity.

1 Introduction

The fundamental concept of Shannon information is uniquely determined by four simple requirements, *continuity, strong additivity, monotonicity*, and a *normalization* (Shannon, 1948). Continuity implies that small perturbations of the underlying probability distribution have only small effects on the information measure, and this is of course very appealing. Strong additivity refers to the requirement that the chain rule $H(ZY) = H(Y) + H(Z|Y)$ holds. Similar conditions are also satisfied, mutatis mutandis, for the derived concepts of conditional and mutual information, as well as for other information measures, such as interaction information/co-information (McGill, 1954; Bell, 2003) or total correlation/multi-information (Watanabe, 1960; Studenỳ and Vejnarová, 1998).

Williams and Beer (2010) proposed to decompose the mutual information that several random variables $Y_1, \dots, Y_k$ have about a target variable S into various components that quantify how much information these variables possess individually, how much they share and how much they need to combine to become useful. That is, one wants to disentangle how information about S is distributed over the $Y_1, \dots, Y_k$. Again, various requirements can be imposed, with varying degrees of plausibility, upon such a decomposition. There are several candidate solutions, and not all of them satisfy all those requirements. Curiously, however, previous considerations did not include continuity and strong additivity. While Bertschinger et al. (2013) did consider chain rule-type properties, none of the information measures defined within the context of information decompositions satisfies any of these chain rule properties (Rauh et al., 2014).

In this contribution, we evaluate which of the various proposed decompositions satisfy *continuity* and *additivity*. Here, additivity (without strong) is required only for independent variables (see Definition 6 below). Additivity (together with other properties) may replace strong additivity when defining Shannon information axiomatically (see Csiszár 2008 for an overview). The importance of additivity is also discussed by Matveev and Portegies (2017).

We consider the case where all random variables are finite, and we restrict to the bivariate case $k = 2$. A bivariate information decomposition consists of three functions SI, UI and CI that depend on the joint distribution of three variables S, Y, Z, and that satisfy:

$$\begin{aligned} I(S;YZ) &= \underbrace{SI(S;Y,Z)}_{\text{shared}} + \underbrace{CI(S;Y,Z)}_{\text{complementary}} + \underbrace{UI(S;Y\backslash Z)}_{\text{unique } (Y \text{ wrt } Z)} + \underbrace{UI(S;Z\backslash Y)}_{\text{unique } (Z \text{ wrt } Y)}, \\ I(S;Y) &= SI(S;Y,Z) + UI(S;Y\backslash Z), \quad I(S;Z) = SI(S;Y,Z) + UI(S;Z\backslash Y). \end{aligned} \tag{1}$$

Hence, $I(S;YZ)$ is decomposed into a shared part that is contained in both Y and Z, a complementary (or synergistic) part that is only available from (Y, Z) together, and unique parts contained exclusively in either Y or Z.

To define a bivariate information decomposition in this sense, it suffices to define either of SI, UI or CI. The other functions are then determined from (1). The linear system (1) consists of three equations in four unknowns, where the two unknowns $UI(S;Y\backslash Z)$ and $UI(S;Z\backslash Y)$ are related. Thus, when starting with a function UI to define an information decomposition, the following consistency condition must be satisfied:

$$I(S;Y) + UI(S;Z\backslash Y) = I(S;Z) + UI(S;Y\backslash Z). \tag{2}$$

If consistency is not given, one may try to adjust the proposed measure of unique information to enforce consistency using a construction from Banerjee et al. (2018) (see Section 2).

As mentioned above, several bivariate information decompositions have been proposed (see Section 2 for a list). However, there are still holes in our understanding of the properties of those decompositions that have been proposed so far. This paper investigates the continuity and additivity properties of some of these decompositions.

Continuity is understood with respect to the canonical topology of the set of joint distributions of finite variables of fixed sizes. When P_n is a sequence of joint distributions with $P_n \to P$, does $SI_{P_n}(S;Y,Z) \to SI_P(S;Y,Z)$? Most, but not all, proposed information decompositions are continuous (i.e. SI, UI and CI are all continuous). If an information decomposition is continuous, one may ask whether it is differentiable, at least at probability distributions of full support. Among the information decompositions that we consider, only the decomposition I_{IG} (Niu and Quinn, 2019) is differentiable. Continuity and smoothness are discussed in detail in Section 3.

The second property that we focus on is additivity, by which we mean that SI, UI and CI behave additively when a system can be decomposed into (marginally) independent subsystems (see Definition 6 in Section 4). This property corresponds to the notion of *extensivity* as used in thermodynamics. Only the information decomposition I_{BROJA} (Bertschinger et al., 2014) in our list satisfies this property. A weak form of additivity, the *identity axiom* proposed by Harder et al. (2013), is well-studied and is satisfied by other bivariate information decompositions.

2 Proposed information decompositions

We now list the bivariate information decompositions that we want to investigate. The last paragraph mentions further related information measures. We denote information decompositions by I, with sub- or superscripts. The corresponding measures SI, UI and CI inherit these decorations.

We use the following notation: S, Y, Z are random variables with finite state spaces $\mathcal{S}$, $\mathcal{Y}$, $\mathcal{Z}$. The set of all probability distributions on a set $\mathcal{X}$ (i.e. the probability simplex over $\mathcal{X}$) is denoted by $\mathbb{P}_{\mathcal{X}}$. The joint distribution P of S, Y, Z is then an element of $\mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$.

- $\boldsymbol{I_{\min}}$ Together with the information decomposition framework, Williams and Beer (2010) also defined an information decomposition $I_{\min}$. Let

$$I(S=s;Y) = \sum_{y\in\mathcal{Y}} P(y|s)\log\frac{P(s|y)}{P(s)}, \quad I(S=s;Z) = \sum_{z\in\mathcal{Z}} P(z|s)\log\frac{P(s|z)}{P(s)}$$

be the *specific information* of the outcome $S=s$ about Y and Z, respectively. Then

$$SI_{\min}(S;Y,Z) = \sum_{s\in\mathcal{S}} P(s)\min\big\{I(S=s;Y), I(S=s;Z)\big\}.$$

$I_{\min}$ has been criticized, because it assigns relatively large values of shared information, conflating "the same amount of information" with "the same information" (Harder et al., 2013; Griffith and Koch, 2014).

- $\boldsymbol{I_{\mathrm{MMI}}}$ A related information decomposition is given by

$$SI_{\mathrm{MMI}}(S;Y,Z) = \min\big\{I(S;Y), I(S;Z)\big\}.$$

Even more severely than $I_{\min}$, this information decomposition conflates "the same amount of information" with "the same information." Still, formally, this definition produces a valid bivariate information decomposition and thus serves as a useful benchmark. The axioms imply that $SI(S;Y,Z) \le SI_{\mathrm{MMI}}(S;Y,Z)$ for any other bivariate information decomposition. For multivariate Gaussian variables, many information decompositions actually agree with I_{MMI} (Barrett, 2015).

- $\boldsymbol{I_{\mathrm{red}}}$ To address the criticism of $I_{\min}$, Harder et al. (2013) introduced a bivariate information decomposition as follows. Let $\mathcal{Z}' := \{z \in \mathcal{Z} : P(Z=z) > 0\}$ be the support of Z, and let

$$
\begin{aligned}
P_{S|y\searrow Z} &= \arg\min_{Q\in\mathrm{conv}\{P(S|z):z\in\mathcal{Z}'\}} D(P(S|y)\|Q),\\
I_S(y\searrow Z) &= D(P(S|y)\|P(S)) - D(P(S|y)\|P_{S|y\searrow Z}),\\
I_S(Y\searrow Z) &= \textstyle\sum_{y\in\mathcal{Y}} P(y) I_S(y\searrow Z).
\end{aligned}
$$

Then

$$
SI_{\mathrm{red}}(S;Y,Z) = \min\big\{I_S(Y\searrow Z), I_S(Z\searrow Y)\big\}.
$$

- $\boldsymbol{I_{\mathrm{BROJA}}}$ Motivated from decision theoretic considerations, Bertschinger et al. (2014) introduced the bivariate information decomposition I_{BROJA}. Given $P\in\mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$, let Δ_P denote the set of joint distributions of (Y,Z,S) that have the same marginals on (S,Y) and (S,Z) as P. Then define the unique information that Y conveys about S with respect to Z as

$$
UI_{\mathrm{BROJA}}(S;Y\backslash Z) := \min_{Q\in\Delta_P} I_Q(S;Y|Z).
$$

I_{BROJA} leads to a concept of synergy that agrees with the synergy measure defined by Griffith and Koch (2014).

- $\boldsymbol{I_{\mathrm{dep}}}$ James et al. (2018) define the following bivariate information decomposition: Given the joint distribution $P\in\mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$ of (S,Y,Z), let $P_{Y-S-Z} = P(S,Y)P(S,Z)/P(S)$ be the probability distribution in $\mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$ that maximizes the entropy among all distributions Q with $Q(S,Y) = P(S,Y)$ and $Q(S,Z) = P(S,Z)$. Similarly, let P_Δ be the probability distribution in $\mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}}$ that maximizes the entropy among all distributions Q with $Q(S,Y) = P(S,Y)$ and $Q(S,Z) = P(S,Z)$ and $Q(Y,Z) = P(Y,Z)$ (unlike for P_{Y-S-Z}, no explicit formula for P_Δ can be given). Then

$$
UI_{\mathrm{dep}}(S;Y\setminus Z) = \min\big\{I_{P_{Y-S-Z}}(S;Y|Z), I_{P_\Delta}(S;Y|Z)\big\}.
$$

This definition is motivated in terms of a lattice of all sensible marginal constraints when maximizing the entropy, as in the definition of P_{Y-S-Z} and P_Δ (see James et al. 2018 for the details).

- $\boldsymbol{I_\cap^*}$, $\boldsymbol{I_\cap^\wedge}$ **and** $\boldsymbol{I_\cap^{\mathrm{GH}}}$ The information decompositions $I_\cap^\wedge$ (Griffith et al., 2014), $I_\cap^{\mathrm{GH}}$ (Griffith and Ho, 2015) and $I_\cap^*$ (Kolchinsky, 2019) present three different approaches to try to represent the shared information in terms of a random variable Q:

$$
\begin{aligned}
SI_\cap^\wedge(S;Y,Z) &= \max\Big\{I(Q;S) : Q = f(Y) = f'(Z)\text{ a.s.}\Big\},\\
SI_\cap^{\mathrm{GH}}(S;Y,Z) &= \max\Big\{I(Q;S) : I(S;Q|Y) = I(S;Q|Z) = 0\Big\},
\end{aligned}
$$

$$SI_{\cap}^{*}(S;Y,Z) = \max\Big\{I(Q;S) : P(s,q) = \sum_{y} P(s,y)\lambda_{q|y} = \sum_{z} P(s,z)\lambda'_{q|z}\Big\},$$

where the optimization runs over all pairs of (deterministic) functions f, f' (for $SI_{\cap}^{\wedge}$), all joint distributions of four random variables S, X, Y, Q that extend the joint distribution of S, X, Y (for $SI_{\cap}^{\mathrm{GH}}$), and all pairs of stochastic matrices $\lambda_{q|y}, \lambda'_{q|z}$ (for $SI_{\cap}^{*}$), respectively. One can show that $SI_{\cap}^{\wedge}(S;Y,Z) \leq SI_{\cap}^{\mathrm{GH}}(S;Y,Z) \leq SI_{\cap}^{*}(S;Y,Z)$ (Kolchinsky, 2019).

The $I_{\cap}^{*}$-decomposition draws motivation from considerations of channel preorders, in a similar spirit as Banerjee et al. (2018), and it is related to ideas from Bertschinger and Rauh (2014). Kolchinsky (2019) shows that there is a deep analogy between $I_{\cap}^{*}$ and I_{BROJA}.

• $\boldsymbol{I_{\mathrm{IG}}}$ Niu and Quinn (2019) presented a bivariate information decomposition I_{IG} based on information geometric ideas. While their construction is very elegant, it only works for joint distributions P of full support (i.e. $P(s,y,z) > 0$ for all s, y, z). It is unknown whether it can be extended meaningfully to all joint distributions. Numerical evidence exists that a unique continuous extension is possible at least to some joint distributions with restricted support (see examples by Niu and Quinn 2019).

For any $t \in \mathbb{R}$, consider the joint distribution

$$P^{(t)}(s,y,z) = \frac{1}{c_t} P^{t}_{S-Y-Z}(s,y,z) P^{1-t}_{S-Z-Y}(s,y,z) = \frac{1}{c_t} P(y,z)P(s|y)^{t}P(s|z)^{1-t},$$

where c_t is a normalizing constant, and let $P^{*} = \arg\min_{t\in\mathbb{R}} D(P\|P^{(t)})$. Then

$$SI_{\mathrm{IG}}(S;Y,Z) = D(P\|P^{*}), \quad UI_{\mathrm{IG}}(S;Y\setminus Z) = D(P^{*}\|P_{S-Z-Y}).$$

• **The UI construction** Given an information measure that captures some aspect of unique information but that fails to satisfy the consistency condition (2), one may construct a corresponding bivariate information decomposition as follows:

Lemma 1. *Let $\delta : \mathbb{P}_{\mathcal{S}\times\mathcal{Y}\times\mathcal{Z}} \to \mathbb{R}$ be a non-negative function that satisfies*

$$\delta(S;Y\setminus Z) \leq \min\{I(S;Y), I(S;Y|Z)\}.$$

Then a bivariate information decomposition is given by

$$\begin{aligned}
UI_{\delta}(S;Y\setminus Z) &= \max\big\{\delta(S;Y\setminus Z), \delta(S;Z\setminus Y) + I(S;Y) - I(S;Z)\big\},\\
UI_{\delta}(S;Z\setminus Y) &= \max\big\{\delta(S;Z\setminus Y), \delta(S;Y\setminus Z) + I(S;Z) - I(S;Y)\big\},\\
SI_{\delta}(S;Z,Y) &= \min\big\{I(S;Y) - \delta(S;Y\setminus Z), I(S;Z) - \delta(S;Z\setminus Y)\big\},\\
CI_{\delta}(S;Z,Y) &= \min\big\{I(S;Y|Z) - \delta(S;Y\setminus Z), I(S;Z|Y) - \delta(S;Z\setminus Y)\big\}.
\end{aligned}$$

Proof. The proof follows just as the proof of Banerjee et al. (2018, Proposition 13). □

We call the construction of Lemma 1 the *UI construction.* The unique information UI_δ returned by the UI construction is the smallest UI-function of any bivariate information decomposition with $UI \geq \delta$.

As Banerjee et al. (2018) show, the decomposition I_{red} is an example of this construction. As another example, as Banerjee et al. (2018) and Rauh et al. (2019) suggested, the UI construction can be used to obtain bivariate information decompositions from the one- or two-way secret key rates and related information functions that have been defined as bounds on the secret key rates, such as the intrinsic information (Maurer and Wolf, 1997), the reduced intrinsic information (Renner and Wolf, 2003), or the minimum intrisic information (Gohari and Anantharam, 2010).

• **Other decompositions** Several other measures have been proposed that are motivated by the framework of Williams and Beer (2010) but that leave the framework. Ince (2017) defines a decomposition I_{ccs}, which satisfies (1), but in which SI_{ccs}, UI_{ccs} and CI_{ccs} may take negative values. The SPAM decomposition of (Finn and Lizier, 2018) consists of non-negative information measures that decompose the mutual information, but this decomposition has a different structure, with alternating signs and twice as many terms. Both approaches construct “pointwise” decompositions, in the sense that SI, UI and CI can be naturally expressed as expectations, in a similar way that entropy and mutual information can be written as expectations (see Finn and Lizier 2018 for details).

Since these measures do not lie in our direct focus, we omit their definitions. Nevertheless, one can ask the same questions: Are the corresponding information measures continuous, and are they additive? For the constructions in Finn and Lizier (2018), both continuity and additivity (as a consequence of a chain rule) are actually postulated. On the other hand, I_{ccs} is neither continuous (as can be seen from its definition) nor additive (since it does not satisfy the identity property).

3 Continuity

Most of the information decompositions that we consider are continuous. Moreover, the UI construction preserves continuity: if δ is continuous, then UI_δ is continuous. The notable exceptions to continuity are I_{red} and the $I_\cap$ decompositions (see Lemmas 2 and 4 below). For SI_{red}, this is due to its definition in terms of conditional probabilities. Thus, SI_{red} is continuous when restricted to probability distributions of full support. For $SI_\cap^*$, discontinuities also appear for sequences $P_n \to P$ where the support does not change. For the SI_{IG} information decomposition, one should keep in mind that it is only defined for probability distributions with full support. It is currently unknown whether it can be continuously extended to probability distributions.

Clearly, continuity is a desirable property, but is it essential? A discontinuous information measure might still be useful, if the discontinuity is not too severe. For example, the Gács-Körner common information $C(Y \wedge Z)$ (Gács and Körner, 1973) is an information measure that vanishes except on a set of measure zero. Clearly, such an information measure is difficult to estimate. The $I_\cap$ decompositions are related to $C(Y \wedge Z)$, and so their discontinuity is almost as severe (see Lemma 4). On the other hand, the I_{red}-decomposition is continuous at distributions of full support. If the discontinuity is well-behaved and well understood, then such a decomposition may still be useful for certain applications. Still, a discontinuous information decomposition challenges the intuition, and any discontinuity must be interpreted (such as the discontinuity of $C(Y \wedge Z)$ can be explained and interpreted (Gács and Körner, 1973)).

If an information decomposition is continuous, one may ask whether it is differentiable, at least at probability distributions of full support. For almost all information decompositions that we consider, the answer is no. This is easy to see for those information decompositions that involve a minimum of finitely many smooth functions ($SI_{\min}$, SI_{MMI}, SI_{red}, SI_{dep}). For SI_{BROJA}, we refer to Rauh and Schünemann (2021). Only SI_{IG} is differentiable for distributions of full support[1].

Two further related properties that have been defined for information measures are asymptotic continuity and locking. As Rauh et al. (2019) show, I_{BROJA} satisfies both properties. For the other information decompositions, it is not known.

Lemma 2. *SI_{red} is not continuous.*

Proof. $I_S(Y \searrow Z)$ and $I_S(Z \searrow Y)$ are defined in terms of conditional probability $P(S|Y = y)$ and $P(S|Z = z)$, which are only defined for those y, z with $P(Y = y) > 0$ and $P(Z = z) > 0$. Therefore, $I_S(Y \searrow Z)$ and $I_S(Z \searrow Y)$ are discontinuous when probabilities tend to zero. □

A concrete example is given below.

Example 3 (SI_{red} is not continuous)**.** *For $0 \le a \le 1$, suppose that the joint distribution of S, Y, Z has the following marginal distributions:*

s	y	$P_a(s,y)$
1	0	$\frac{a}{2}$
1	1	$\frac{1}{2} - \frac{a}{2}$
0	1	$\frac{1}{4}$
0	2	$\frac{1}{4}$

s	z	$P_a(s,z)$
0	0	$\frac{a}{2}$
0	1	$\frac{1}{2} - \frac{a}{2}$
1	1	$\frac{1}{4}$
1	2	$\frac{1}{4}$

.

Observe the symmetry of Y, Z. For $a > 0$, the conditional distributions of S given Y and Z are, respectively:

y	$P_a(S\|y)$
0	$\binom{0}{1}$
1	$\binom{\frac{1}{3-2a}}{\frac{2-2a}{3-2a}}$
2	$\binom{1}{0}$

and

z	$P_a(S\|z)$
0	$\binom{1}{0}$
1	$\binom{\frac{2-2a}{3-2a}}{\frac{1}{3-2a}}$
2	$\binom{0}{1}$

.

[1]Personal communication with the authors Niu and Quinn (2019).

Therefore, $I_S(Y \searrow Z) = I(S;Y) = I(S;Z) = I_S(Z \searrow Y)$.

For $a = 0$*, the conditional distributions* $P(S|Y = 0)$ *and* $P(S|Z = 0)$ *are not defined. It follows that* $I_S(Y \searrow Z) = I_S(Z \searrow Y) < I(S;Y) = I(S;Z)$*. In total,*

$$\lim_{a\to 0+} SI_{\text{red}}(S;Y,Z) = \lim_{a\to 0+} I(S;Y) > I_S(Y \searrow Z) = SI_{\text{red}}(S;Y,Z).$$

Lemma 4. $SI_\cap^*$*,* $I_\cap^\wedge$ *and* $I_\cap^{\text{GH}}$ *are not continuous.*

Proof. By Kolchinsky (2019, Section V.B and Theorem 5), for all three measures, $SI_\cap(YZ;Y,Z)$ equals the Gács-Körner common information $C(Y \wedge Z)$ (Gács and Körner, 1973), which is not continuous. □

A concrete example is given below.

Example 5 ($SI_\cap^*$ is not continuous)**.** *Suppose that the joint distribution of* S, Y, Z *has the following marginal distributions, for* $-1 \le a \le 1$*:*

s	y	$P_a(s,y)$
0	0	$\frac{1}{3}$
1	0	$\frac{1}{6} - \frac{a}{6}$
1	1	$\frac{1}{6} + \frac{a}{6}$
2	1	$\frac{1}{3}$

s	z	$P_a(s,z)$
0	0	$\frac{1}{3}$
1	0	$\frac{1}{6}$
1	1	$\frac{1}{6}$
2	1	$\frac{1}{3}$

.

For $a = 0$*, the marginal distributions of the pairs* (S, Y) *and* (S, Z) *are identical, whence* $SI_\cap^*(S;Y,Z) = I(S;Y) = I(S;Z)$.

Now let $a \ne 0$*. According to the definition of* $SI_\cap^*$*, we need to find stochastic matrices* $\lambda_{q|y}, \lambda'_{q|z}$ *that satisfy the condition*

$$P(s,q) = \sum_y P(s,y)\lambda_{q|y} = \sum_z P(s,z)\lambda'_{q|z} \tag{3}$$

For $s = 0$ *and* $s = 2$*, this condition implies* $\lambda_{q|0} = \lambda'_{q|0}$ *and* $\lambda_{q|1} = \lambda'_{q|1}$*. For* $s = 1$*, the same condition gives* $a(\lambda_{q|1} - \lambda_{q|0}) = 0$*. In the case* $a \ne 0$*, this implies that* $\lambda_{q|1} = \lambda_{q|0}$ *and that* Q *is independent of* S*. Therefore,* $SI_\cap^*(S;Y,Z) = 0$ *for* $a = 0$*.*

4 Additivity

Definition 6. *An information measure* $I(X_1, \dots, X_n)$ *(i.e. a function of the joint distribution of n random variables) is* additive *if and only if the following holds: If* $(X_1, \dots, X_n)$ *are independent of* $(Y_1, \dots, Y_n)$*, then*

$$I(X_1Y_1, X_2Y_2, \dots, X_nY_n) = I(X_1, \dots, X_n) + I(Y_1, \dots, Y_n).$$

The information measure is superadditive*, if, under the same assumptions,*

$$I(X_1Y_1, X_2Y_2, \dots, X_nY_n) \ge I(X_1, \dots, X_n) + I(Y_1, \dots, Y_n).$$

Among those decompositions that we consider, only one is additive:

Lemma 7. I_{BROJA} *is additive.*

Proof. This is Bertschinger et al. (2014, Lemma 19). □

All other information decompositions that we consider are not additive. However, in all information decompositions that we consider, SI is superadditive and UI is subadditive (Theorem 10).

Again, additivity is a desirable property, but is it essential? As in the case of continuity, we argue that non-additivity challenges the intuition, and any non-additivity must be interpreted. Why is it plausible that the shared information contained in two independent pairs is more than the sum of the individual shared informations, and how can one explain that the unique information is subadditive?

A related weaker property is additivity under i.i.d. sequences, i.e. when, in the definition of additivity, (S_1, Y_1, Z_1) and (S_2, Y_2, Z_2) are identically distributed. One can show that I_{red}, I_{MMI}, I_{dep} and I_{IG} (and, of course, I_{BROJA}) are additive under i.i.d. sequences, but not $I_{\min}$. The UI construction gives additivity under i.i.d. sequences I_δ if δ is additive under i.i.d. sequences. The proof of these statements is similar to the proof for additivity (given below) and omitted. For the $I_\cap$ decompositions, it is not as easy to see, and so we currently do not know whether additivity under i.i.d. sequences holds.

Lemma 8. *1. If I_1 and I_2 are superadditive, then* $\min\{I_1, I_2\}$ *is superadditive.*

2. If, in addition, there exist distributions P, Q with $I_1(P) < I_2(P)$ and $I_1(Q) > I_2(Q)$, then $\min\{I_1, I_2\}$ *is not additive.*

Proof. 1. With $X_1, \dots, X_n, Y_1, \dots, Y_n$ as in the definition of superadditivity,

$$\begin{aligned}
\min\big\{&I_1(X_1Y_1, X_2Y_2, \dots, X_nY_n), I_2(X_1Y_1, X_2Y_2, \dots, X_nY_n)\big\} \\
&\geq \min\big\{I_1(X_1, \dots, X_n) + I_1(Y_1, \dots, Y_n), I_2(X_1, \dots, X_n) + I_2(Y_1, \dots, Y_n)\big\} \\
&\geq \min\big\{I_1(X_1, \dots, X_n), I_2(X_1, \dots, X_n)\big\} + \min\big\{I_1(Y_1, \dots, Y_n), I_2(Y_1, \dots, Y_n)\big\}.
\end{aligned}$$

2. In this inequality, if $X_1, \dots, X_n \sim P$ and $Y_1, \dots, Y_n \sim Q$, then the right hand side equals $I_1(X_1, \dots, X_n) + I_2(Y_1, \dots, Y_n)$, which makes the inequality strict. □

As a consequence:

Lemma 9. *If δ is subadditive, then UI_δ is subadditive, SI_δ is superadditive, but neither is additive.*

Theorem 10. *$I_{\min}$, I_{MMI}, I_{red}, I_{dep}, I_{IG} as well as the $I_\cap$ decompositions are superadditive, but not additive.*

Proof. For I_{MMI}, the claim follows directly from Lemma 8. The same is true for I_{dep}, since $I_{P_{Y-S-Z}}(S; Y|Z)$ and $I_{P_\Delta}(S; Y|Z)$ are additive, and also for I_{red}, since $I_S(Y \searrow Z)$ and

$I_S(Z \searrow Y)$ are superadditive. For $I_{\min}$, the same argument as in the proof of Lemma 8 applies, since the specific information is additive, in the sense that

$$I(S_1S_2 = s_1s_2; Y_1Y_2) = I(S_1 = s_1; Y_1) + I(S_2 = s_2; Y_2). \qquad \square$$

Next, consider SI_{IG}. For $i = 1, 2$

$$P_i^{(t)}(s_i, y_i, z_i) = \frac{1}{c_{i,t}} P(y_i, z_i) P(s_i|y_i)^t P(s_i|z_i)^{1-t}.$$

Then

$$P^{(t)}(s_1s_2, y_1y_2, z_1z_2) = P_1^{(t)}(s_1, y_1, z_1) P_2^{(t)}(s_2, y_2, z_2)$$

and

$$D(P\|P^{(t)}) = D(P_1\|P_1^{(t)}) + D(P_2\|P_2^{(t)}),$$

where P_i denotes the marginal distribution of S_i, Y_i, Z_i for $i = 1, 2$. It follows that

$$\begin{aligned} SI(S_1S_2; Y_1Y_2, Z_1Z_2) = \min_{t\in\mathbb{R}} D(P\|P^{(t)}) &\geq \min_{t\in\mathbb{R}} D(P_1\|P_1^{(t)}) + \min_{t\in\mathbb{R}} D(P_2\|P_2^{(t)}) \\ &= SI(S_1; Y_1, Z_1) + SI(S_2; Y_2, Z_2). \end{aligned}$$

If $\arg\min_{t\in\mathbb{R}} D(P_1\|P_1^{(t)}) \neq \arg\min_{t\in\mathbb{R}} D(P_2\|P_2^{(t)})$, then strict inequality holds.

Theorem 11. *$I_\wedge$, I_{GH} and I_* are additive.*

Proof. First, consider $I_\wedge$. As Griffith et al. (2014) show, $SI_\wedge(S_1S_2; Y_1Y_2, Z_1Z_2) = I(S_1S_2; Q)$, where Q is the *common random variable* (Gács and Körner, 1973), which satisfies $Q = Q_1Q_2$, where Q_j is the common random variable of Y_j and Z_j. Therefore,

$$\begin{aligned} SI_\wedge(S_1S_2; Y_1Y_2, Z_1Z_2) =& I(S_1S_2; Q_1Q_2) = I(S_1; Q_1) + I(S_2; Q_2) \\ =& SI_\wedge(S_1; Y_1, Z_1) + SI_\wedge(S_2; Y_2, Z_2). \end{aligned}$$

Let $SI_\cap \in \{SI_{\mathrm{GH}}, SI_*\}$. To see that $SI_\cap$ is superadditive, suppose that $SI_\cap(S_j; Y_j, Z_j) = I(Q_j; S_j)$. The joint distribution of S_1, S_2, Q_1, Q_2 defined by $P(s_1s_2q_1q_2) = P(s_1q_1)P(s_2q_2)$ is feasible for the optimization problem in the definition of $SI_\cap(S_1S_2; Y_1Y_2, Z_1Z_2)$. Therefore,

$$\begin{aligned} SI_\cap(S_1S_2; Y_1Y_2, Z_1Z_2) \geq I(S_1S_2; Q_1Q_2) &= I(S_1; Q_1) + I(S_2; Q_2) \\ &= SI_\cap(S_1; Y_1, Z_1) + SI_\cap(S_2; Y_2, Z_2). \end{aligned}$$

To prove subadditivity of I_{GH}, let Q be as in the definition of $SI_{\mathrm{GH}}(S_1S_2; Y_1Y_2, Z_1Z_2)$, with $S_1, S_2, Y_1, Y_2, Z_1, Z_2$ as in Definition 6. The chain rule implies $I(S_1S_2; Q) = I(S_1; Q) + I(S_2; Q|S_1)$, where $I(S_2; Q|S_1) = \sum_{s_1} P(S_1 = s_1) I(S_2; Q|S_1 = s_1)$. Choose s_1^* such that $I(S_2; Q|S_1 = s_1^*) = \max_{s_1} I(S_2; Q|S_1 = s_1)$.

Construct two random variables Q_1, Q_2 as follows: Q_1 is independent of S_2, Y_2, Z_2 and satisfies $P(Q_1|S_1, Y_1, Z_1) = P(Q|S_1, Y_1, Z_1)$. Q_2 is independent of S_1, Y_1, Z_1 and satisfies $P(Q_2|S_2, Y_2, Z_2) = P(Q|S_2, Y_2, Z_2, S_1 = s_1^*)$. By construction, Q_1Q_2 is independent of

S_1S_2 given Y_1Y_2, and Q_1Q_2 is independent of S_1S_2 given Z_1Z_2. The statement follows from

$$\begin{aligned}
&SI_{\mathrm{GH}}(S_1S_2;Y_1Y_2,Z_1Z_2)+SI_{\mathrm{GH}}(S_1S_2;Y_1Y_2,Z_1Z_2)\\
&\qquad\geq I(S_1S_2;Q_1Q_2)=I(S_1;Q_1)+I(S_2;Q_2)=I(S_1;Q)+I(S_2;Q|S_1=s_1)\\
&\qquad\geq I(S_1;Q)+I(S_2;Q|S_1)=I(S_1S_2;Q)=SI_{\mathrm{GH}}(S_1S_2;Y_1Y_2,Z_1Z_2).
\end{aligned}$$

To prove subadditivity for I_*, we claim that for all random variables S, Y, Z there exist random variables S', Y', Z' with $P(S,Y) = P(S',Y')$, $P(S,Z) = P(S',Z')$ and $I_*(S;Y,Z) = I_{\mathrm{GH}}(S';Y',Z') = I_*(S;Y,Z)$. This correspondence can be chosen such that $(S_1S_2)' = S_1'S_2'$, $(Y_1Y_2)' = Y_1'Y_2'$ and $(Z_1Z_2)' = Z_1'Z_2'$, where $S_1'Y_1'Z_1'$ is independent of $S_2'Y_2'Z_2'$. Thus,

$$\begin{aligned}
SI_*(S_1S_2;Y_1Y_2,Z_1Z_2)&=SI_{\mathrm{GH}}\big((S_1S_2)';(Y_1Y_2)',(Z_1Z_2)'\big)\\
&=SI_{\mathrm{GH}}(S_1';Y_1',Z_1')+SI_{\mathrm{GH}}(S_2';Y_2',Z_2')\\
&\leq SI_*(S_1';Y_1',Z_1')+SI_*(S_2';Y_2',Z_2').
\end{aligned}$$

To prove the claim, suppose that $SI_*(S;Y,Z) = I(S;Q)$, with Q as in the definition of SI_*. Define random variables S', Y', Z', Q' such that

$$P(S'Y'Z'Q'=syzq)=P(SQ=sq)P(Y=y|SQ=sq)P(Z=z|SQ=sq).$$

Then $P(S'Y' = sy) = P(SY = sy)$ and $P(S'Z' = sz) = P(SZ = sz)$. Since SI_* only depends on the (SY) and (SZ)-marginals, $SI_*(S;Y,Z) = SI_*(S';Y',Z')$. Moreover,

$$SI_*(S;Y,Z)=I(S;Q)=I(S';Q')\leq SI_{\mathrm{GH}}(S';Y',Z')\leq SI_*(S';Y',Z').$$

The claim follows from this. □

5 Conclusions

We have studied measures that have been defined for bivariate information decompositions, asking whether they are continuous and/or additive. The only information decomposition that is both continuous and additive is I_{BROJA}.

While there are many continuous information decompositions, it seems difficult to construct differentiable information: Currently, the only differentiable example is I_{IG} (which, however, is only defined in the interior of the probability simplex). It would be interesting to know which other smoothness properties are satisfied by the proposed information decompositions, such as locking and asymptotic continuity.

It also seems to be difficult to construct additive information decompositions, with I_{BROJA} being the only known example. In contrast, many known information decompositions are additive under i.i.d. sequences. In the other direction, it would be worthwhile to have another look at stronger versions of additivity, such as chain rule-type properties. Bertschinger et al. (2013) concluded that such chain rules prevent a straightforward extension of decompositions to the non-bivariate case along the lines of Williams and Beer (2010). It has recently been argued (e.g. Rauh 2017) that a general information decomposition likely needs a structure that differs from the proposal by Williams and Beer (2010), whence another look at chain rules may be worthwhile.

Acknowledgement

PB and GM have been supported by the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 757983).

References

P. K. Banerjee, E. Olbrich, J. Jost, and J. Rauh. Unique informations and deficiencies. In *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 32–38, 2018.

A. B. Barrett. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Phys. Rev. E*, 91:052802, 2015.

A. J. Bell. The co-information lattice. In *Proc. Fourth Int. Symp. Independent Component Analysis and Blind Signal Separation (ICA 03)*, 2003.

N. Bertschinger and J. Rauh. The Blackwell relation defines no lattice. In *IEEE International Symposium on Information Theory*, pages 2479–2483, 2014.

N. Bertschinger, J. Rauh, E. Olbrich, and J. Jost. Shared information — new insights and problems in decomposing information in complex systems. In *Proc. ECCS 2012*, pages 251–269. Springer, 2013.

N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014.

I. Csiszár. Axiomatic characterizations of information measures. *Entropy*, 10:261–273, 2008.

C. Finn and J. T. Lizier. Pointwise partial information decomposition using the specificity and ambiguity lattices. *Entropy*, 20(4), 2018.

P. Gács and J. Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973.

A. A. Gohari and V. Anantharam. Information-theoretic key agreement of multiple terminals-Part I. *IEEE Transactions on Information Theory*, 56(8):3973–3996, 2010.

V. Griffith and T. Ho. Quantifying redundant information in predicting a target random variable. *Entropy*, 17(7):4644–4653, 2015.

V. Griffith and C. Koch. Quantifying synergistic mutual information. In M. Prokopenko, editor, *Guided Self-Organization: Inception*, volume 9, pages 159–190. Springer Berlin Heidelberg, 2014. doi: 10.1007/978-3-642-53734-9_6.

V. Griffith, E. K. P. Chong, R. G. James, C. J. Ellison, and J. P. Crutchfield. Intersection information based on common randomness. *Entropy*, 16(4):1985–2000, 2014.

M. Harder, C. Salge, and D. Polani. A bivariate measure of redundant information. *Phys. Rev. E*, 87:012130, Jan 2013.

R. Ince. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy*, 19(7):318, 2017.

R. James, J. Emenheiser, and J. Crutchfield. Unique information via dependency constraints. *Journal of Physics A*, 52(1):014002, 2018.

A. Kolchinsky. A novel approach to multivariate redundancy and synergy. *arXiv:1908.08642*, 2019.

R. Matveev and J. W. Portegies. Tropical limits of probability spaces, Part I: The intrinsic Kolmogorov-Sinai distance and the asymptotic equipartition property for configurations. *arXiv:1704.00297*, 2017.

U. Maurer and S. Wolf. The intrinsic conditional mutual information and perfect secrecy. In *Proc. IEEE ISIT*, 1997. doi: 10.1109/ISIT.1997.613003.

W. McGill. Multivariate information transmission. *IRE Trans. Inf. Theory*, 4(4):93–111, 1954. ISSN 2168-2690. doi: 10.1109/TIT.1954.1057469.

X. Niu and C. Quinn. A measure of synergy, redundancy, and unique information using information geometry. In *Proc. IEEE ISIT*, 2019.

J. Rauh. Secret sharing and shared information. *Entropy*, 19(11):601, 2017.

J. Rauh and M. Schünemann. Properties of unique information. *Kybernetika*, 57:383–403, 2021.

J. Rauh, N. Bertschinger, E. Olbrich, and J. Jost. Reconsidering unique information: Towards a multivariate information decomposition. In *Proc. IEEE ISIT*, pages 2232–2236, 2014.

J. Rauh, P. K. Banerjee, E. Olbrich, and J. Jost. Unique information and secret key decompositions. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 3042–3046, 2019.

R. Renner and S. Wolf. New bounds in secret-key agreement: The gap between formation and secrecy extraction. In *Advances in Cryptology - EUROCRYPT 2003, Warsaw, Poland*, pages 562–577, 2003.

C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, 1948.

M. Studený and J. Vejnarová. The multiinformation function as a tool for measuring stochastic dependence. In *Learning in graphical models*, pages 261–297. Springer, 1998.

S. Watanabe. Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, 4(1):66–82, 1960.

P. Williams and R. Beer. Nonnegative decomposition of multivariate information. *arXiv:1004.2515v1*, 2010.

Probabilistic representation of spatial fuzzy sets

Lubomír Soukup[1]

[1]Institute of Information Theory and Automation of the Czech Academy of Sciences
soukup@utia.cas.cz

Abstract

Membership function of a given fuzzy set is expressed by probability that a point belongs in the fuzzy set. Such a membership function is derived from probability distribution of points on the boundary of the fuzzy set. Polygonal boundary is considered. Spatial operations (conjunction, disjunction, complement) are defined accordingly. Several application areas are mentioned, namely classification of land cover, cadastral mapping, material quality analysis, interferometric monitoring of bridges.

1 Introduction

Several attempts had been made to represent uncertain real objects by means of precise mathematical tools. Two most successful approaches – probability theory and fuzzy sets theory – yet seem to be non-compatible. These two approaches have been heavily applied to plenty of real-world problems, but still there are little understanding of their mutual relationship. The early attempt to bring together probability and fuzziness was made by the founder of the fuzzy sets theory (Zadeh, 1968). His achievements were further worked out from more general point of view in (Singpurwalla and Jane M. Booker, 2004). Researchers who have tried making fuzzy set theory and probability theory work in concert usually agree with (Zadeh, 1995) that the two approaches are complementary rather than compatible or competitive. These authors customarily conclude that vaugness and randomness demonstrate two different aspects of uncertainty of the real world. Therefore fuzzy sets differ from imprecise regions, membership function differs from probability measure. Overview of such results is presented in (Schmitz and Morris, 2006).

Nevertheless, there are some cases that allow membership function of a fuzzy set to gain probabilistic meaning. Passing reference of this possibility can be found in book (Viertl, 1996). This case typically occures when fuzzy sets are spatially defined, namely as geographic regions. Some interesting attempts to represent imprecise regions stem from geographical information sciences, e. g. (Cunha and Martins, 2014), (Bruin, 2000). Therefore, geographical motivation stands beyond the approach addressed in this contribution.

2 Formulation of the problem

We are searching for a spatial fuzzy region whose boundary is uncertain due to imprecise position of points on its boundary. Provided that the boundary of the region has a polygonal shape the problem can be concisely formulated as follows:

2.1 Given:

1. two–dimensional closed polygonal region with imprecise vertices,
2. probability distribution of each vertex of the polygonal boundary.

2.2 Required:

1. probability that a point in 2D plane belongs to the given region,
2. fuzzy set whose membership function is defined by the probabilities evaluated by means of 1.,
3. fuzzy set operations (conjunction, disjunction, complement) of probabilistic fuzzy sets created by means of 2.

The given polygonal region can look as Figure 1 shows.

3 Solution — 1D case

3.1 Membership function

Principle of creating the probabilistic membership function can be easily explained in 1D case. Let us suppose that the spatial fuzzy set is formed by a line segment $\mathcal{F}_{A,B}$. Boundary of this set consists of two imprecise points, position of which is given by random variables X_A, X_B. Probability that some fixed point x_U belongs in the fuzzy set $\mathcal{F}_{A,B}$ can be evaluated by

$$\begin{aligned} P(x_U \in \mathcal{F}_{A,B}) &= P((X_A < x_U) \wedge (X_B > x_U) \,|\, X_A < X_B) = \\ &= \frac{\displaystyle\int_{-\infty}^{x_U} f_A(x)\,dx \int_{x_U}^{\infty} f_B(y)\,dy}{\displaystyle\int_{-\infty}^{\infty} f_B(y) \int_{-\infty}^{y} f_A(x)\,dx\,dy} \end{aligned} \tag{1}$$

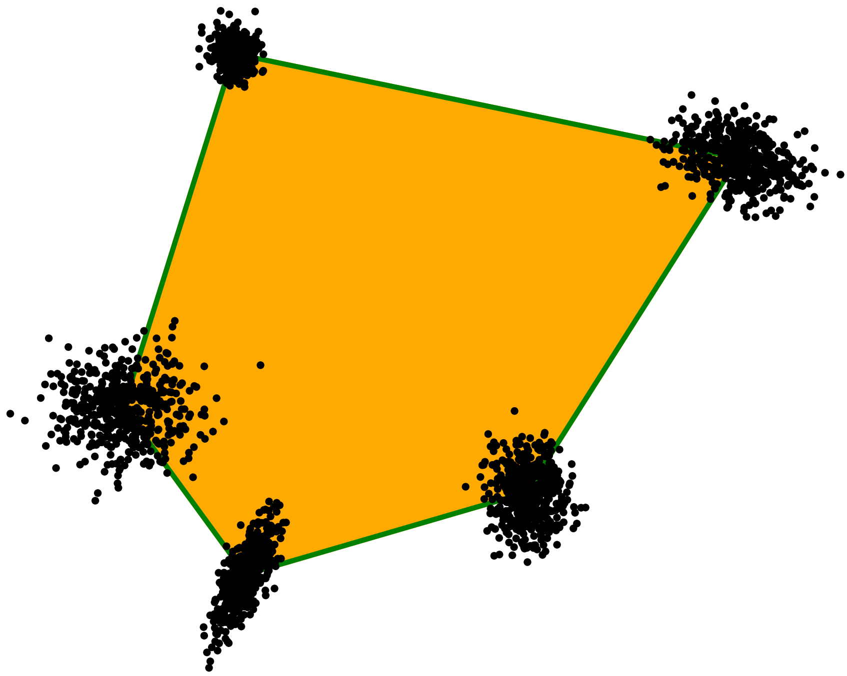


Figure 1: Polygonal region with imprecise vertices. Precision of each vertex is modeled by a 2D probability distribution (black clusters).

Function f_A, resp. f_B stands for probability density function of random variable X_A, resp. X_B. If the probability density functions are Gaussian $\mathcal{N}(\hat{x}_A, \sigma_A)$, resp. $\mathcal{N}(\hat{x}_B, \sigma_B)$, the resulting probability is quite simple:

$$P(x_U \in \mathcal{F}_{A,B}) = \frac{\left(1 + \operatorname{erf}\left(\frac{x_U - \hat{x}_A}{\sqrt{2}\,\sigma_A}\right)\right)\left(1 - \operatorname{erf}\left(\frac{x_U - \hat{x}_B}{\sqrt{2}\,\sigma_B}\right)\right)}{2\left(1 + \operatorname{erf}\left(\frac{\hat{x}_B - \hat{x}_A}{\sqrt{2\,(\sigma_A^2 + \sigma_B^2)}}\right)\right)} \ . \tag{2}$$

Function erf stands for error function, i. e.

$$\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt \ .$$

Membership function of fuzzy set $\mathcal{F}_{A,B}$ can therefore be defined by probability (1).

$$\mu_{\mathcal{F}_{A,B}} \ : \ \mathbb{R} \to \langle 0, 1 \rangle \ : \ \xi \mapsto \mu_{\mathcal{F}_{A,B}}(\xi) := P(\xi \in \mathcal{F}_{A,B}) \ . \tag{3}$$

Illustration of the probabilistic membership function gives Figure 2.

3.2 Fuzzy set operations

3.2.1 Probabilistic fuzzy conjunction

Probability that some fixed point x_U belongs in conjunction of two fuzzy sets $\mathcal{F}_{A,B}$, $\mathcal{F}_{C,D}$ can be expressed similarly as in (1).

$$\begin{aligned} & P(x_U \in \mathcal{F}_{A,B} \wedge \mathcal{F}_{C,D}) = \\ = \; & P((X_A < x_U < X_B) \wedge (X_C < x_U < X_D) \,|\, (X_A < X_B) \wedge (X_C < X_D)) = \\ = \; & \frac{\int_{-\infty}^{x_U} f_A(w)\,dw \int_{x_U}^{\infty} f_B(x)\,dx \int_{-\infty}^{x_U} f_C(y)\,dy \int_{x_U}^{\infty} f_D(z)\,dz}{\int_{-\infty}^{\infty} f_B(x) \int_{-\infty}^{x} f_A(w)\,dw\,dx \int_{-\infty}^{\infty} f_D(z) \int_{-\infty}^{z} f_C(y)\,dy\,dz} \,. \end{aligned} \tag{4}$$

If the border probability density functions are Gaussian $\mathcal{N}(\hat{x}_A, \sigma_A)$, $\mathcal{N}(\hat{x}_B, \sigma_B)$, $\mathcal{N}(\hat{x}_C, \sigma_C)$, $\mathcal{N}(\hat{x}_D, \sigma_D)$, the resulting probability will be as follows:

$$\begin{aligned} & P(x_U \in \mathcal{F}_{A,B} \wedge \mathcal{F}_{C,D}) = \\ = \; & \frac{\left(1 + \operatorname{erf}\left(\frac{x_U - \hat{x}_A}{\sqrt{2}\,\sigma_A}\right)\right)\left(1 - \operatorname{erf}\left(\frac{x_U - \hat{x}_B}{\sqrt{2}\,\sigma_B}\right)\right)\left(1 + \operatorname{erf}\left(\frac{x_U - \hat{x}_C}{\sqrt{2}\,\sigma_C}\right)\right)\left(1 - \operatorname{erf}\left(\frac{x_U - \hat{x}_D}{\sqrt{2}\,\sigma_D}\right)\right)}{4\left(1 + \operatorname{erf}\left(\frac{\hat{x}_B - \hat{x}_A}{\sqrt{2\,(\sigma_A^2 + \sigma_B^2)}}\right)\right)\left(1 + \operatorname{erf}\left(\frac{\hat{x}_D - \hat{x}_C}{\sqrt{2\,(\sigma_C^2 + \sigma_D^2)}}\right)\right)} \,. \end{aligned} \tag{5}$$

3.2.2 Probabilistic fuzzy disjunction

Probabilistic fuzzy disjunction can be easily deduced from conjunction operation (4) with aid of elementary theorem of probability theory

$$P(K \vee L) = P(K) + P(L) - P(K \wedge L) \tag{6}$$

which holds for any random events K, L.

$$P(x_U \in \mathcal{F}_{A,B} \vee \mathcal{F}_{C,D}) = P(x_U \in \mathcal{F}_{A,B}) + P(x_U \in \mathcal{F}_{C,D}) - P(x_U \in \mathcal{F}_{A,B} \wedge \mathcal{F}_{C,D}) \tag{7}$$

Evaluation of (7) under assumption of normal distribution can be simply done with substitutions (2), (5).

3.2.3 Probabilistic fuzzy complement

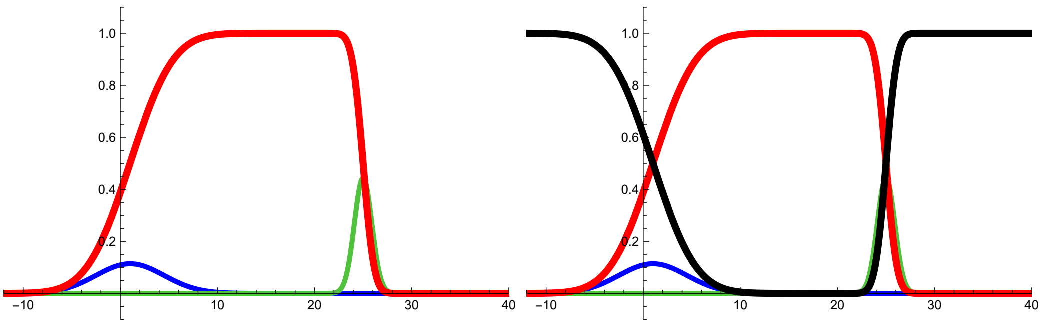
Membership function of probabilistic fuzzy complement can be obtained from (1) as easy as fuzzy conjunction from (4). The same theorem (6) can be used with special option

$L = \neg K$. Due to this option, $P(K \vee \neg K) = 1$ and $P(K \wedge \neg K) = 0$ for any random event K. Theorem (6) then claims

$$1 = P(K) + P(\neg K) \;.$$

Thus the following equalities hold for fuzzy complement $\neg\mathcal{F}_{A,B}$.

$$P(x_U \in \neg\mathcal{F}_{A,B}) = P(x_U \notin \mathcal{F}_{A,B}) = 1 - P(x_U \in \mathcal{F}_{A,B}) \tag{8}$$



$\mathcal{F}_{A,B}$ $\neg\mathcal{F}_{A,B}$

Figure 2: Membership function of probabilistic fuzzy set $\mathcal{F}_{A,B}$ – drawn by red line. Border probability density functions f_A (blue line) and f_B (green line) are shown on the left and right hand side. The right part of the figure shows also membership function of probabilistic fuzzy complement $\neg\mathcal{F}_{A,B}$ (black line).

3.2.4 Membership functions of the fuzzy set operations

Membership functions of the above introduced fuzzy set operations are given by

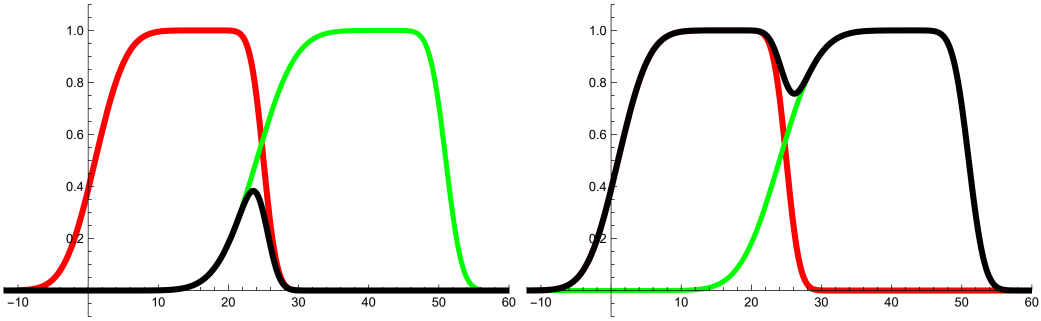
$$\begin{aligned} \mu_{\mathcal{F}_{A,B}\wedge\mathcal{F}_{C,D}}(\xi) &:= P(\xi \in \mathcal{F}_{A,B} \wedge \mathcal{F}_{C,D}) \;, \\ \mu_{\mathcal{F}_{A,B}\vee\mathcal{F}_{C,D}}(\xi) &:= P(\xi \in \mathcal{F}_{A,B} \vee \mathcal{F}_{C,D}) = \mu_{\mathcal{F}_{A,B}}(\xi) + \mu_{\mathcal{F}_{C,D}}(\xi) - P(\xi \in \mathcal{F}_{A,B} \wedge \mathcal{F}_{C,D}) \;, \\ \mu_{\neg\mathcal{F}_{A,B}}(\xi) &:= P(\xi \notin \mathcal{F}_{A,B}) = 1 - \mu_{\mathcal{F}_{A,B}}(\xi) \;. \end{aligned} \tag{9}$$

These membership functions are defined by means of the source membership functions $\mu_{\mathcal{F}_{A,B}}$, $\mu_{\mathcal{F}_{C,D}}$ except fuzzy conjunction. Therefore, proper definiton of fuzzy conjunction has to be accomplished by the following two–step procedure.

1. extract border density functions f_A, f_B from $\mu_{\mathcal{F}_{A,B}}$ and f_C, f_D from $\mu_{\mathcal{F}_{C,D}}$,
2. evaluate $P(x_U \in \mathcal{F}_{A,B} \wedge \mathcal{F}_{C,D})$ with the aid of (4) .

The first step of this procedure may not be satisfactorily achievable since the membership functions $\mu_{\mathcal{F}_{A,B}}$, $\mu_{\mathcal{F}_{C,D}}$ can be given in other ways than by (1). Extraction of border density functions from an arbitrary membership function is subject of further research.

Illustration of the fuzzy set operations gives Figure 3 and Figure 2.



$\mathcal{F}_{A,B} \wedge \mathcal{F}_{C,D}$ $\mathcal{F}_{A,B} \vee \mathcal{F}_{C,D}$

Figure 3: Membership functions of probabilistic fuzzy conjunction $\mathcal{F}_{A,B} \wedge \mathcal{F}_{C,D}$ (left) and disjunction $\mathcal{F}_{A,B} \vee \mathcal{F}_{C,D}$ (right) are drawn by black line. The original fuzzy sets are shown by red and light green colors.

4 Solution — 2D case

4.1 2D membership function

Two–dimensional generalization of formula (1) becomes much more complicated if a polygonal boundary is considered. Closed polygon with n vertices is given. Coordinates of i-th vertex are given in form of column vector $\mathbf{x}_i \in \mathbb{R}^2$, $i \in \{1, 2, \ldots, n\}$. Sides of the polygon must not intersect each other if they are not adjacent. Moreover, for the sake of simplicity, the polygon is supposed to be convex. Coordinates of the all vertices creates $2n$-dimensional vector

$$\mathbf{x} := [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \ .$$

Area of the given polygon is

$$S(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{n} \det([\mathbf{x}_i, \mathbf{x}_{i+1}]) \ , \tag{10}$$

where $[\mathbf{x}_i, \mathbf{x}_{i+1}]$ is 2×2 real matrix for each $i \in \{1, 2, \ldots, n\}$,

$$\mathbf{x}_{n+1} := \mathbf{x}_1 \ .$$

Note that $S(\mathbf{x}) \geq 0$ if the polygon is oriented counter–clockwise.

Let each vertex be a random point in $\mathbb{R}^2$. Then the all random vertices form random vector

$$\mathbf{X} := [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n]$$

that creates random boundary of a polygon $\mathcal{F}_n$. Such imprecise polygon can or cannot cover a fixed point $\mathbf{x}_U \in \mathbb{R}^2$. Probability that the point $\mathbf{x}_U$ lies inside the imprecise polygon is

$$P(\mathbf{x}_U \in \mathcal{F}_n) = P(\mathbf{X} \in \mathcal{U}(\mathbf{x}_U) \,|\, (S(\mathbf{X}) > 0) \wedge \mathcal{C}(\mathbf{X})) \ . \tag{11}$$

Statement $\mathcal{C}(\mathbf{x})$ claims that polygon with vertices $\mathbf{x}$ is convex.

Set $\mathcal{U}(\mathbf{x}_U)$ stands for the all possible convex polygons that include point $\mathbf{x}_U$. It is defined by

$$\mathcal{U}(\mathbf{x}_U) := \{\mathbf{x} \in \mathbb{R}^{2n} \mid \left(\bigwedge_{i=1}^{n} \kappa(\mathbf{x}_i, \mathbf{x}_{i+1}, \mathbf{x}_U) > 0 \right) \wedge \mathcal{C}(\mathbf{x})\} ,$$

where function κ expresses perpendicular oriented distance of the point $\mathbf{x}_U$ from line that passes through oriented tuple of points $\mathbf{a}, \mathbf{b}$.

$$\kappa : \mathbb{R}^6 \to \mathbb{R} \ : \ [\mathbf{a}, \mathbf{b}, \mathbf{x}_U] \mapsto (\mathbf{a} - \mathbf{x}_U) \cdot \frac{(\mathbf{b} - \mathbf{a})^\perp}{||\mathbf{b} - \mathbf{a}||} \tag{12}$$

Symbol $^\perp$ stands for perpendicularity.

$$^\perp : \mathbb{R}^2 \to \mathbb{R}^2 \ : \ [x, y] \mapsto [x, y]^\perp := [-y, x] .$$

If the vertices of the polygonal boundary have probability distribution with density function f, then

$$P(\mathbf{x}_U \in \mathcal{F}_n) = \frac{\underbrace{\int\int \cdots \int\int}_{\mathcal{U}(\mathbf{x}_U)} f(\mathbf{x})\, d\mathbf{x}}{P((S(\mathbf{X}) > 0) \wedge \mathcal{C}(\mathbf{X}))} . \tag{13}$$

The multiple integral is $2n$-tuple.

$$P((S(\mathbf{X}) > 0) \wedge \mathcal{C}(\mathbf{X})) = \underbrace{\int\int \cdots \int\int}_{(S(\mathbf{x})>0) \wedge \mathcal{C}(\mathbf{x})} f(\mathbf{x})\, d\mathbf{x} .$$

4.2 2D fuzzy set operations

4.2.1 Probabilistic 2D fuzzy conjunction

Conjunction of two 2D fuzzy sets, say $\mathcal{F}_n$, $\mathcal{G}_m$ can be defined similarly as in 1D case (4) by probability

$$P(\mathbf{x}_U \in \mathcal{F}_n \wedge \mathcal{G}_m) = \frac{\underbrace{\int\int \cdots \int\int}_{\mathcal{U}(\mathbf{x}_U) \wedge \mathcal{V}(\mathbf{x}_U)} f(\mathbf{x})\, g(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}}{P(S(\mathbf{x}) > 0 \wedge S(\mathbf{y}) > 0 \wedge \mathcal{C}(\mathbf{x}) \wedge \mathcal{C}(\mathbf{y}))} , \tag{14}$$

where $\mathbf{y}$ is $2m$-dimensional vector which contains coordinates of m vertices of a convex polygon. Corresponding random vector $\mathbf{Y}$ has probability density function g.

$$\mathcal{V}(\mathbf{x}_U) := \{\mathbf{y} \in \mathbb{R}^{2m} \mid \left(\bigwedge_{i=1}^{m} \kappa(\mathbf{y}_i, \mathbf{y}_{i+1}, \mathbf{x}_U) > 0 \right) \wedge \mathcal{C}(\mathbf{y})\} ,$$

where function κ was introduced in (12).

$$P(S(\mathbf{X}) > 0 \wedge S(\mathbf{Y}) > 0 \wedge \mathcal{C}(\mathbf{X}) \wedge \mathcal{C}(\mathbf{Y})) = \underbrace{\int\int \cdots\cdots \int\int}_{S(\mathbf{x})>0 \wedge S(\mathbf{y})>0 \wedge \mathcal{C}(\mathbf{x}) \wedge \mathcal{C}(\mathbf{y})} f(\mathbf{x})\, g(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} \; .$$

The above multiple integrals are $2(n+m)$-tuple.

Suitable approximation have to be applied to evaluate the multiple integration for normal densities f, g.

4.2.2 Probabilistic 2D fuzzy disjunction

$$P(x_U \in \mathcal{F}_n \vee \mathcal{G}_m) = P(x_U \in \mathcal{F}_n) + P(x_U \in \mathcal{G}_m) - P(x_U \in \mathcal{F}_n \wedge \mathcal{G}_m) \; . \qquad (15)$$

4.2.3 Probabilistic 2D fuzzy complement

$$P(x_U \in \neg\mathcal{F}_n) = P(x_U \notin \mathcal{F}_n) = 1 - P(x_U \in \mathcal{F}_n) \; . \qquad (16)$$

4.2.4 Membership functions of the 2D fuzzy set operations

Membership functions of the 2D fuzzy set operations are given similarly as in (9) by

$$\begin{aligned} \mu_{\mathcal{F}_n \wedge \mathcal{G}_m}(\xi) &:= P(\xi \in \mathcal{F}_n \wedge \mathcal{G}_m) \; , \\ \mu_{\mathcal{F}_n \vee \mathcal{G}_m}(\xi) &:= P(\xi \in \mathcal{F}_n \vee \mathcal{G}_m) = \mu_{\mathcal{F}_n}(\xi) + \mu_{\mathcal{G}_m}(\xi) - P(\xi \in \mathcal{F}_n \wedge \mathcal{G}_m) \; , \\ \mu_{\neg\mathcal{F}_n}(\xi) &:= P(\xi \notin \mathcal{F}_n) = 1 - \mu_{\mathcal{F}_n}(\xi) \; . \end{aligned} \qquad (17)$$

Problem of extracting probability density functions f, g from membership functions $\mu_{\mathcal{F}_n}$, $\mu_{\mathcal{G}_m}$ is much more arduous in 2D than in 1D case. Solution of this problem has crucial importance in real examples, namely in cartography and material analysis. These examples will be addressed in the next section.

5 Applications

Probabilistic fuzzy regions can be found all around. Therefore, the designed approach has many practical applications, namely classification of land cover, cadastral mapping, material quality analysis, interferometric monitoring of bridges.

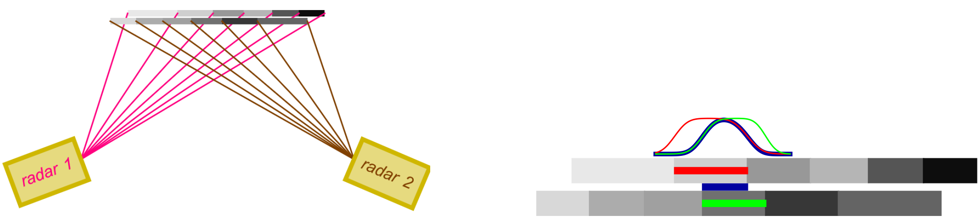
5.1 Cartography and material analysis

One of the most frequent task of digital cartography is classification of land cover. Simplest case of the classification is recognition of certain region of interest against other type of earth surface. The region of interest has to be localized by determination of points on its boundary and input into geographic information system (GIS). Precision of these points can be inferred from so called probability map which is by–product of classification procedure. The resulting region of interest then will be obtained as an imprecise region similar to Figure 1. Spatial operations that are necessary component of every GIS can then be realized by the fuzzy operation designed in this contribution. Similar problems with imprecise boundary occures in cadastral systems where parcels have polygonal shapes.

Another application area has been emerged in material quality analysis. For example, shape and size changes of microscopic grains in cocnrete blocks under radiation exposure are important for security assessment of nuclear powerplants. Size of these grains can be precisely estimated with the aid of the probabilistic representation designed in this contribution.

5.2 Radar interferometry

Radar interferometry is very effective method for determination of descends and rises of a bridge under traffic load. Magnitude and direction of movements of the bridge body can be very precisely (up to 0.01 mm) measured by two ground based radars (GB-RAR), see Figure 4. Unfortunately, part of the bridge body (so called range–bin) that corresponds to the measurement cannot be determined precisely. Gray–scale rectangles on Figure 4 represent response of radar rays from different range–bins. Regions of the range–bins are imprecise, so that their conjunctions are imprecise as well. Probabilistic quantification of these imprecisions which is offered in the presented fuzzy approach can improve interpolation quality of the bridge movement in an arbitrary point.



bridge monitoring　　　fuzzy region on the bridge

Figure 4: **Bridge monitoring** (left) by simultaneous measurement of two interferometric radars. Gray–scale rectangles represent response of radar rays from different range–bins. **Fuzzy region on the bridge** (right) shows conjunction of two overlapping range–bins (red and green). The conjunction is marked by dark blue color.

6 Conclusion

Representation of a fuzzy set of polygonal shape was designed in such a way that probability of membership of any point in the set is derived from imprecision of the polygonal boundary. This representation enables a non-traditional definition of fuzzy set operations (conjunction, disjunction, complement) that produces other fuzzy sets which gain the same probabilistic interpretation. Several real-world applications of the presented approach were addressed.

Acknowledgement

The research was supported by the Ministry of Interior of the Czech Republic in the framework of grant project "The advanced technology of rapid determination of bridges deformation by radar interferometry and its use in diagnostics", No. VI20192022167.

References

S. D. Bruin. Querying probabilistic land cover data using fuzzy set theory. *International Journal of Geographical Information Science*, 14(4):359–372, 2000.

E. Cunha and B. Martins. Using one-class classifiers and multiple kernel learning for defining imprecise geographic regions. *International Journal of Geographical Information Science*, 28(11):2220–2241, 2014.

A. Schmitz and A. Morris. Modelling and manipulating fuzzy regions: strategies to define the topological relations between fuzzy regions. *Control and Cybernetics*, 35:245–261, 2006.

N. D. Singpurwalla and Jane M. Booker. Membership Functions and Probability Measures of Fuzzy Sets. *Journal of the American Statistical Association*, 99(467):867–877, 2004.

R. Viertl. *Statistical Methods for Non-Precise Data.* CRC Press, 1996.

L. A. Zadeh. Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23:421–427, 1968.

L. A. Zadeh. Discussion: Probability Theory and Fuzzy Logic Are Complementary Rather than Competitive. *Technometrics*, 37:271–276, 1995.

Modeling COVID Pandemics: Strengths and Weaknesses of Epidemic Models

Martin Šmíd[1]

[1]Department of Econometrics, Institute of Information Theory and Automation of the Czech Academy of Sciences
smid@utia.cas.cz

Abstract

We generally discuss modeling the present COVID pandemics. We argue that useful models have to be simple in the first case, yet their uncertainty has to be handled properly. In order to study circumstances of the upcoming wave of infection, we construct a simple stochastic model and present predictions it gives. We conclude that the autumn wave is most likely unavoidable and suggest concentrating to mitigation.

1 Introduction

Vast majority of epidemic models are derived from the seminal SIR model (Kermack and McKendrick, 1927). These models are all both *explainable* (opposed to black boxes, the mechanism of their predictions is understandable by humans) and *interpretable* (the assumed causes produce expected effects, coherent with common sense as well as scientific state of the art); as Rudin (2019) correctly points out, especially the latter property is important whenever the model takes part in decisions on "high-stake" matters, which a pandemic certainly is. Moreover, SIR-like models comply with the well known Hill's criteria of causation (Šmíd and Kuběna, 2022; Hill, 1965). Despite this, their application during the recent covid pandemics became subject of numerous controversies, mostly due to untreated or wrongly interpreted uncertainty, wider sense co-linearity, model risk and the models' normativeness, resulting e.g. in the prevention paradox.

Apparently, uncertainty is the most severe limiting factor of epidemic models as well as the most common source of misunderstanding. Uncertainty can be either about the characteristics of the disease and their variations, about human behavior, about efficiency of counter-epidemic measures, immunity waning, or the rate of reporting. Unfortunately, these uncertainties multiply in time, which brings problems especially to forecasting in the phase of the epidemic growth.

Moreover, parameters of the models often fail to be (in the statistical sense) identified. Unfortunately, often the parameters evaluating impacts of individual counter-epidemic measures suffer from this problem, so it is virtually impossible to quantify the measures' efficiency.

Further, there is a risk of omitted or a falsely considered factors in the models, called *model risk*. The omitted factors could be e.g. seasonality or the onset of a new virus variant. The false factor could, on the other hand, be an alleged effect of otherwise ineffective measure. Due to the multiplicative nature of the models, this can totally invalidate forecasts stemming from them.

Another problem is that epidemic models are potentially normative, meaning that people can act to prevent the predicted effects, e.g. when, in fear of the forecast, they impose measures and/or start behaving protectively, which consequently leads to damping of the epidemics contrary to the prediction; this effect is called the prevention paradox.

Yet these shortcomings are severe, they need not prevent the quantitative models from being used. If the uncertainty is correctly taken into account and the models are wisely formulated, then the models can give reasonable forecasts with a reasonable uncertainty. The co-linearity may be handled as well, yet for the price of not modeling all the influences individually. The model risk can be, to certain extent, guarded by means of suitable performance measure, which should, however, also evaluate the quantification of forecasts' uncertainty (see e.g. Bracher et al. (2021)). Finally, the normativeness of the models should be clearly communicated and, instead of unavoidable forecasts, the scenarios should be published, e.g. what would happen if no reaction takes place, what happens if a lockdown is imposed, etc.

In this slightly non-traditional, little informal paper, we demonstrate the ideas mentioned above by constructing a simple model predicting the autumn 2022 wave of Covid infection from the perspective of May 2022.

2 Simplicity First

After two years engagement in quantitative analysis of the COVID pandemic related data, I understood that one of the greatest virtues of successful models is simplicity. Together with Rudin (2019) I argue that added value of complex models in comparison to simple ones is sometimes negligible, especially when the overall uncertainty is high. Simple models, on the other hand, are sooner to be developed, easier to be maintained, easier to be explained, less time-consuming to be dealt with both by its author and the audience, and, most importantly, easier to be intuitively understood by the author, which fact brings him necessary confidence when using, presenting and justifying it.

Of course, there is always a trade-off between complexity and explainability of the model (Gilpin et al., 2018); however, benefits of simple models often prevail in my opinion, yet the price for the simplicity is the fact that we often have to create them ad-hoc for specific situations rather than trying to construct an universal reusable model. My experience can serve as an example: during the first year of pandemic, we developed a complex compartment stochastic SEIR model (Šmíd et al., 2021). Yet it appears to perform well, adjusting it so that it could answer a specific question always took substantial effort as well as time, both human and computational. Moreover, all the (possibly unrelated) parameters had to be always estimated, bringing additional uncertainty, and the evaluation and sensitivity analysis took so much time that it often could not be done. The simple model I am going to present here, on the other hand, has been created out of scratch within a single working day and is implemented using a OpenOffice spreadsheet[1] which means that any computation or parameter change is nearly immediate.

3 When the Next Wave Comes?

The question the intended model is supposed to answer is what will be the autumn COVID wave be like, and whether and how it could be influenced. As we cannot know how the virus will mutate, we shall assume that the currently variant Omicron will keep prevailing. For our analysis, we shall use a discretized and perturbed SIR model:

$$X_t = \rho s_{t-1} c_{t-2} I_{t-1} X_{t-1} + E_t, \qquad t \geq 1$$

where X_t is the number of reported cases of the Omicron variant, ρ is an estimated constant reflecting the infectiousness of the variant, I_t is the ratio of susceptible population, s_t is the seasonal factor, c_t is the risk contact reduction, and E_t is the error term. Two things are important here: the weekly time step and heteroskedasticity.

As for the former: Yet daily data are available, they show significant weekly seasonality with unpredictably changing pattern, modeling of which is a perfect example of an unnecessary complexity bringing unnecessary obstacles and little benefit (the estimate of ρ could be more precise only in case that we can handle the seasonality precisely).

The heteroskedasticity reflects the fact that variance of the errors scales with the cases numbers. In toy textbook models, the cases number would be Poisson, so the scale would be proportional to $\sqrt{X_t}$; in practice, however, the distribution is over-dispersed (Endo et al., 2020; Getz et al., 2006) so the error variance scales rather with X_t. As a consequence, we can reformulate the model as

$$D_t = \rho s_{t-1} c_{t-2} I_{t-1} + e_t, \qquad D_t = \frac{X_t}{X_{t-1}}, \qquad t \geq 1,$$

with e_t being white noise.

The seasonality term we assume to be

$$s_t = 1 + \kappa \cos(\phi + \psi t), \qquad t \geq 1,$$

where ϕ and ψ are such that the maximum of s happens in the middle of January each year. We set $\kappa = 0.18$, which is the value obtained by Šmíd (2022), roughly equal to the following from Gavenčiak et al. (2021).

The contact reduction is measured by the longitudinal study PAQ research (2021); the fact that the epidemic growth depends on c two weeks earlier is discussed in Šmíd and

[1] See `https://github.com/cyberklezmer/epidata/blob/main/autumn22.ods`.

Kuběna (2022); Šmíd (2022) as well as in Šmíd et al. (2021); it should be said, however, that, contrary to the previous years, only little or no contact reduction takes place this year and is not likely to happen in the future, so we keep c in the model only to be able to model potential crisis scenarios.

In line with Šmíd (2022), we further assume that

$$I_t = V_t J_t, \qquad V_t = 1 - \frac{Y_t}{p}, \qquad J_t = 1 - \frac{Z_t}{p}$$

where Y_t and Z_t are the numbers of individuals having the vaccine-induced immunity, post-infection immunity, respectively, and p is the total population of the Czech Republic. In determining Y_t and Z_t, we use quite precious estimates of vaccine effectiveness, post-infection protection and their waning, obtained by our recent work Šmíd et al. (2022). As for vaccination, we take

$$Y_t = Y_t^f + Y_t^b, \qquad Y_t^f = (1 - w_f)Y_{t-1}^f + e_f F_t - \gamma_f B_t, \qquad Y_t^b = (1 - w_b)Y_{t-1}^b + e_b B_t;$$

where F_t and B_t are numbers of newly fully vaccinated, having obtained the booster, respectively, $e_f = 0.45$ and $e_b = 0.61$ are the initial effectiveness of full vaccination, booster, respectively, $w_f = 0.056$ and $w_b = 0.082$ are the weekly rates of waning of the vaccination effectiveness, booster effectiveness, respectively, and γ_f is the rate of Y_t and its hypothetical counterpart with $e_f = 1$, $w_f = 0$.

As for the post-infection immunity, we divide the infected into those, who were infected once by the Omicron variant, those infected once by the other variants, and those who were reinfected by the Omicron or the other variants:

$$Z_t = Z_t^o + Z_t^\delta + Z_t^{o+} + Z_t^{\delta+}.$$

Here,

$$Z_t^{O+} = Z_{t-1}^{O+} + O_t^+, \qquad Z_t^{\delta+} = Z_{t-1}^{\delta+} + \Delta_t^+$$

where O_t^+ and Δ_t^+ are the numbers of new reinfections of those, who were previously infected by the Omicron variant, other variants, respectively, and

$$Z_t^o = (1 - w_o)Z_{t-1}^o + O_t - \gamma_o O_t^+, \qquad Z_t^\delta = (1 - w_\delta)Z_{t-1}^\delta + \Delta_t - \gamma_\delta \Delta_t^+,$$

where O_t and Δ_t are newly coming infections by the Omicron variant, other variants, respectively, w_o and $w_\delta = 0.022$ are waning coefficients, and γ_o and γ_δ are analogous to γ_f. Note that we assume hundred percent initial post-infection immunity which does not wane if the individual has been infected twice. It should be stressed that w_o is still unknown as there is still a short time from the Omicron's emergence so relevant data is still not available. Thus we later perform our analysis for its various values. As not all infections are reported, we assume that

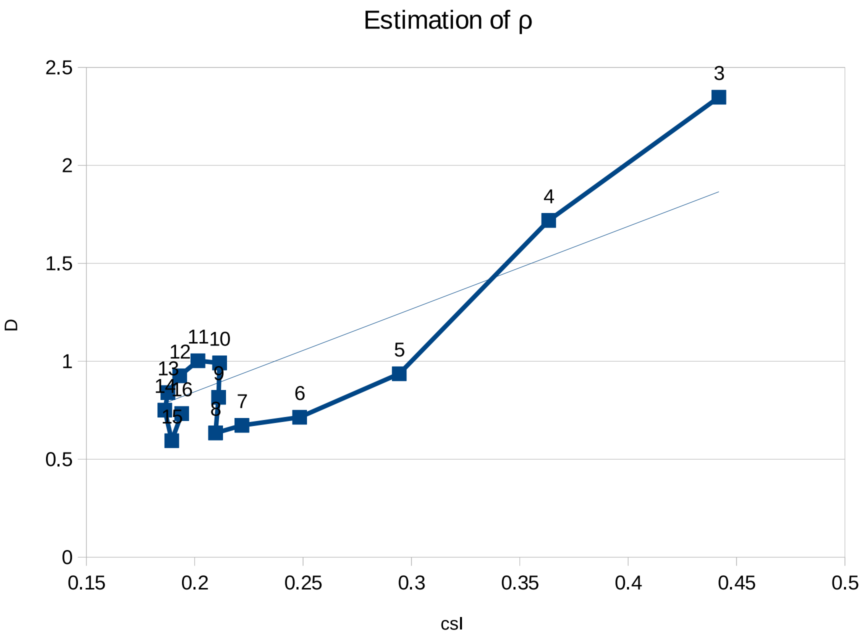
$$O_t = \frac{X_t}{\alpha},$$

where α is the ascertainment rate, i.e. the fraction of reported infections, and we compute Δ_t, O_t^+ and Δ_t^+ from their reported counterparts analogously. In line with Šmíd (2022), we put $\alpha = 0.4$.

4 Results

As inputs of our model, we used publicly available data ČR (2021) and an online variant proportion estimator.[2] For the estimation, we used data from week 2/2022 when the Omicron variant started to prevail, to week 16/2022 – two weeks before the model construction.

The following regression graph with the observation labeled by week numbers shows that the studied dependence may be regarded as linear; however, the onset of more infectious variant BA2 is suggested. We neglect this fact first and use the overall estimate, but we return to this issue later.

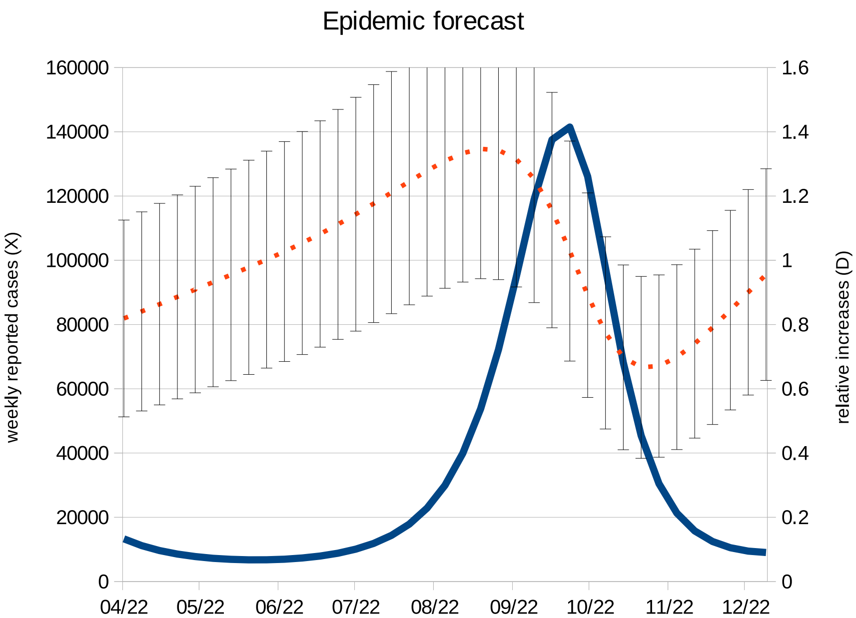


The estimated value of ρ is 4.22(0.25).

Before doing any forecast, we find important to realize that the future behavior of the pandemic depends on many parameters, some of which we are uncertain about, and, yet this additional uncertainty is often difficult to quantify, it has to be added up to the inherent uncertainty, represented by e_t and the estimation error of ρ. Maybe more important, however, is to realize that in systems depending on human behavior, which pandemic certainly is, not only the behavior can change in a reaction to the system (e.g. being careful when infection numbers are high), but it can change also in a reaction to our forecast.

[2] https://covariants.org/

The following graph shows a point forecasts of D and X given that (i) the contact reduction keeps unchanged, being equal to $c_t = 0.9$, (ii) the waning of the post-infection immunity after Omicron infection is the same as that after Delta $w_o = w_\delta$, and (iii) the vaccination rate will not change, i.e. there will be 2000 final doses and 10000 boosters a week:

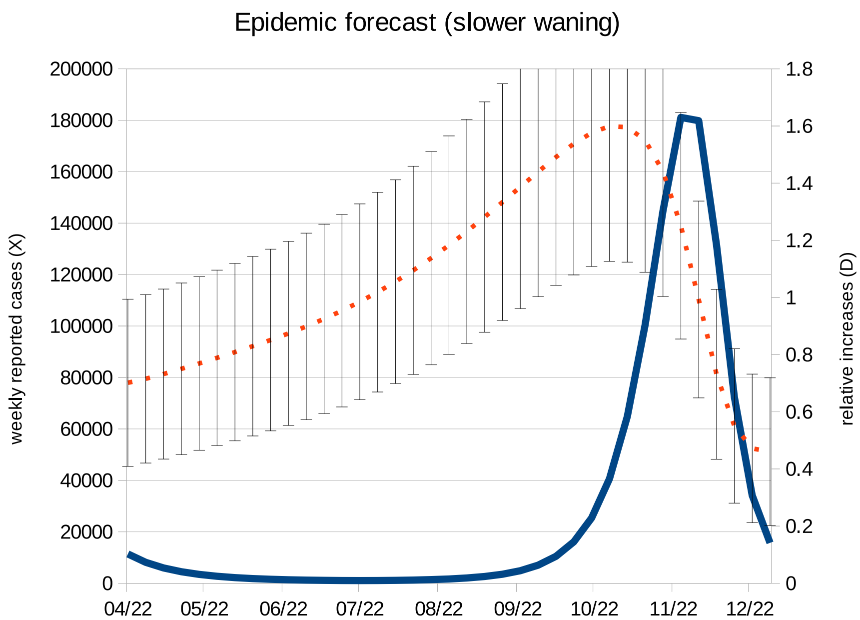


The solid line shows predicted numbers of observed cases X, the dotted line depicts the relative growths D together with a lower estimate of their standard forecast errors – we do not include the part of the uncertainty caused by the fact that I depends on D's, which could, in principle, be evaluated, but this is beyond the scope of this short paper – we only remark that, as the omitted errors would multiply, it is quite clear that the errors explode after the confidence interval for I starts to contain unity. The situation is even more serious for the predictions of X errors of which would start exploding immediately. Thus, taking these as well as the mentioned additional uncertainties into account, it is clear that such a forecast should be interpreted qualitatively rather than quantitatively.

The great uncertainty of the forecast, however, does not mean that the model does not say anything. It is clear, for instance, that D will sooner or later reach unity, because their determinants V and I keep growing (the waning obviously overturns the effects of new vaccinations, new infections, respectively). In this sense, another wave seems unavoidable. Note also that, after the predicted October wave, the forecasts of D quickly approach unity again which means that another wave in the beginning of 2023 is likely.

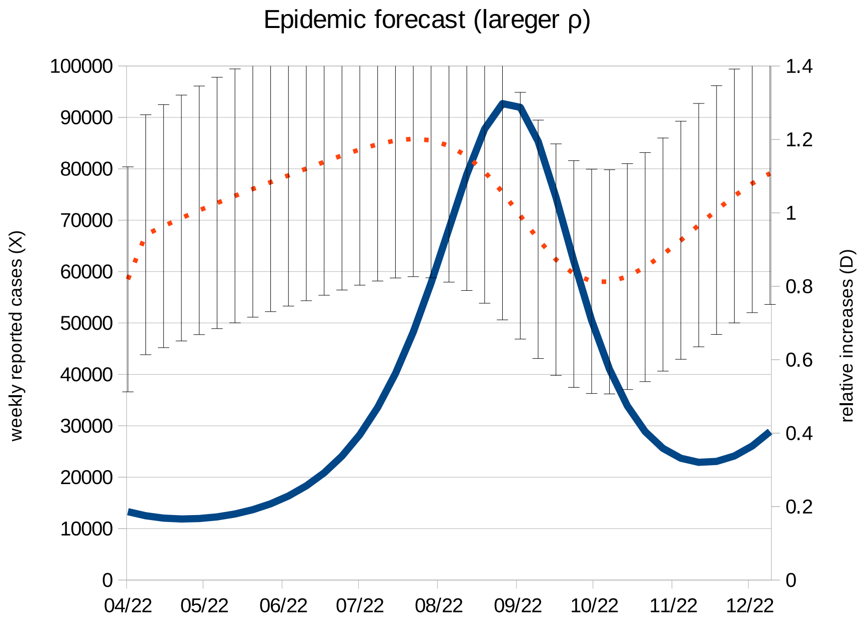
5 Sensitivity Analysis

As it has been already mentioned, two crucial parameters are subject of a great uncertainty. Most important it is the rate w_o of post-infection immunity waning given that the original infection was by Omicron. So far we assumed it to be equal to the same value 0.022 as if the original infection were by other variants. Newly we set it to the waning rate of the immunity against Delta after the Delta infection, i.e. $w_o = 0.003$. The result is following:



The fact that slowing the waning rate seven times only shifts the autumn wave one month later might seem surprising; however, it suffices to realize that there is still more people who were infected by older variants than those who underwent the Omicron infection and that the seasonal component grows in autumn.

The second highly uncertain parameter is ρ. Below is the forecast assuming that, from the time horizon (end of April), ρ starts to greater by 0.5, i.e. $\rho = 4.72$, perhaps due to the fact that a new variant of Omicron prevailed.



Not surprisingly, the expected wave came earlier. It is also less in its peak; however, this does not be a great victory as the numbers start to grow at the end of the year, obviously preparing for the next wave.

6 Discussion and Conclusion

We presented a simple model designed in order to study the circumstances of the expected autumn wave of the COVID infection. Having observed great uncertainty of the model's forecasts, we resorted to qualitative forecasts instead of quantitative ones. Still, however, we dare to conclude that another wave of infection is most likely unavoidable. Question arises, whether it can be averted or at least how it can be mitigated.

The answer to the first question is unfortunately no, the reason being the low effectiveness of the contemporary vaccines against Omicron infection. As we have no other acceptable means of preventing the infection spreading, it remains to take the upcoming wave as a fact and move to mitigation. As Šmíd et al. (2022) show, the existing vaccines are still rather effective against a severe course of the disease, so the most straightforward move is to vaccinate anyone who is or may be endangered. Moreover, still having time enough, it would be reasonable to discuss "logistic" aspects of the wave, namely to prepare measures which would prevent the wave from paralyzing daily life as it nearly happened during the recent wave when e.g. schools could hardly function due to strict quarantine rules.

The presented model, yet simple, can be improved too. As new data come, the waning and effectiveness parameters can be further refined, the impacts on hospitals can be studied and the uncertainty may be quantified more preciously.

References

J. Bracher, E. L. Ray, T. Gneiting, and N. G. Reich. Evaluating epidemic forecasts in an interval format. *PLoS computational biology*, 17(2):e1008618, 2021.

A. Endo et al. Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. *Wellcome Open Research*, 5, 2020.

T. Gavenčiak, J. T. Monrad, G. Leech, M. Sharma, S. Mindermann, J. M. Brauner, S. Bhatt, and J. Kulveit. Seasonal variation in sars-cov-2 transmission in temperate climates. *medRxiv*, 2021.

W. M. Getz, J. O. Lloyd-Smith, P. C. Cross, S. Bar-David, P. L. Johnson, T. C. Porco, and M. S. Sánchez. Modeling the invasion and spread of contagious diseases in heterogeneous populations. *Disease evolution: models, concepts, and data analyses*, 71:113–144, 2006.

L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pages 80–89. IEEE, 2018.

A. B. Hill. The environment and disease: association or causation?, 1965.

W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, 115:700–721, 1927. doi: 10.1098/rspa.1927.0118.

PAQ research. Život během pandemie. `https://zivotbehempandemie.cz/`, 2021. Sociologická studie, cit. 13. 5. 2021.

C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

M. ČR. Covid-19 v Čr: Otevřené datové sady a sady ke stažení. `https://onemocneni-aktualne.mzcr.cz/api/v2/covid-19`, 2021. cit. 22. 5. 2021.

M. Šmíd. *Rok s pandemií covid-19 – reflexe v poločase*, chapter Dosáhneme kolektivní imunity očkováním? Carolinum, 2022. In print. In Czech.

M. Šmíd and A. Kuběna. *Rok s pandemií covid-19 – reflexe v poločase*, chapter Fungují opatření? Korelace versus kauzalita. Carolinum, 2022. In print. In Czech.

M. Šmíd, L. Berec, A. A. Kuběna, R. Levinsky, J. Trnka, V. Tucek, and M. Zajíček. SEIR filter: a stochastic model of epidemics. *medRxiv*, 2021. doi: 10.1101/2021.02.16.21251834.

List of Authors

The 12th Workshop on Uncertainty Processing

June 1–4, 2022

Kutná Hora, Czech Republic

Editors

Milan Studený, Nihat Ay, Giulianella Coletti, Gernot D. Kleiter, Prakash P. Shenoy

Organized by

Institute of Information Theory and Automation, Czech Academy of Sciences

Faculty of Management, Prague University of Economics and Business, Jindřichův Hradec

http://wupes.utia.cas.cz/

matfyzpress

PUBLISHING HOUSE
OF THE FACULTY OF MATHEMATICS AND PHYSICS
CHARLES UNIVERSITY IN PRAGUE

ISBN 978-80-7378-460-7

9 788073 784607