Operationele grafische onzekerheidsmodellen

An Operational Approach to Graphical Uncertainty Modelling

Filip Hermans

Promotor: prof. dr. ir. G. de Cooman
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen: Wiskundige Ingenieurstechnieken

Vakgroep Elektrische Energie, Systemen en Automatisering
Voorzitter: prof. dr. ir. J. Melkebeek
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2011 - 2012

UNIVERSITEIT
GENT

ISBN 978-90-8578-512-5
NUR 916, 984
Wettelijk depot: D/2012/10.500/38

Promotor

prof. dr. ir. Gert de Cooman, UGent

Overige leden van de examencommissie

prof. dr. ir. Dirk Aeyels, UGent
prof. dr. ir. Bernard De Baets, UGent
dr. ir. Erik Quaeghebeur (secretaris), UGent
prof. dr. ir. Alain Sarlette, UGent
prof. dr. ir. Rik Van de Walle (voorzitter), UGent
prof. dr. Paolo Vicig, Università di Trieste
prof. dr. Vladimir Vovk, Royal Holloway, University of London
prof. dr. ir. Joris Walraevens, UGent

Adres

Universiteit Gent
Faculteit Ingenieurswetenschappen
Vakgroep Elektrische Energie, Systemen en Automatisering
Onderzoeksgroep Synthese, Sturen en Modelleren van Systemen

Technologiepark-Zwijnaarde 914
9052 Zwijnaarde
België

Financiering

Dit proefschrift bevat de resultaten van onderzoek gefinancierd door het Bijzonder Onderzoeksfonds.

Contents

Acknowledgments

In true Oscar style (I might even be wearing a costume at the moment you read this), I would like to thank all the people who have helped me make this thesis what it is, and have made the last couple of years of my life the super, happy fun place it has been, and hopefully will remain for a long, long time to come.

For the part of making my thesis to what it is, there are no words that can express my gratitude—besides maybe “thank you very much”—towards my promotor Gert de Cooman. Many thanks for introducing me into the wonderful world of imprecise probabilities and its scientific community. Also many thanks for the patience when going through the many proofs with holes the average Swiss cheese might be jealous of. Also many thanks—and that holds for all of the other members of SYSTeMS—for showing me that not all scientists are in the profitable, but not too environmentally friendly, paper producing business. Most of all, many thanks for being this enjoyable and contagiously enthusiastic person you are.

Most of the comments just made (excluding the Swiss cheese one maybe) actually hold for the other SYSTeMS members too, and therefore, I would like to thank: Keivan, Rolando, Dirk, Mia, Alain, Erik, Filip, Nathan, Frits, Jonathan, Margot, Fortunato, Herman, Nicolae, Arthur, Jasper, Mohammad, Bram, Rene and the Saeco. I would especially like to thank Erik, for being my intellectual big brother who did not learn me to ride my bike, but gave me numerous hints and explanations when I got stuck once more in the most referenced book in this thesis. Another special mention goes to Frits and

Fortunato for the incredibly deep and philosophical discussions we had over lunch and during any other possible break.

I also would like to thank the people from CBD. The bitter-ish taste the science approach in that department gave me, was largely recompensed by the great, zest-of-lemon colleagues (maybe the Mojitos are to blame for the flavour) I encountered there. Thank you: Steven, Naira, Krzysiek, Vladimir, Gert, Jeroen, Elena, Anna, Martin and Erick. Of course I should not forget to mention Cecilia, Luigi, Jeffrey and Stéphane here.

At last I would like to thank my parents, my brothers and Miriam, just for being so important to me.

Gent, May 2012
Filip

Samenvatting
–Summary in Dutch–

De laatste twintig jaar kenden een snelle groei van grafische modellen in artificiële intelligentie en statistiek. Deze modellen combineren grafen en waarschijnlijkheidsrekening om complexe multivariate problemen aan te pakken, en het aantal toepassingen is legio. Zowel in de medische, financiële en biotechnologische wereld alsook in domeinen zoals risico-analyse en defensie. worden deze modellen gebruikt. Omdat de informatie die over een fysisch systeem verkregen wordt vaak imprecies van aard is, blijft het echter een fikse uitdaging om de waarschijnlijkheden in deze modellen te bepalen. Als gevolg hiervan worden vaak te sterke en slecht onderbouwde conclusies getrokken. Het is dan ook niet ongebruikelijk om een sensitiviteitsanalyse uit te voeren. Bij zo'n sensitiviteitsanalyse worden de parameters in het bestudeerde grafische netwerk lichtjes veranderd, en wordt er nagegaan of deze kleine variaties tot dezelfde conclusies leiden. Deze sensitiviteitsanalyse is echter niet altijd afdoende. Het is bijvoorbeeld niet zo dat de overtuigingen van een expert aangaande de mogelijke uitkomsten van een toevallige veranderlijke steeds volledig bepaald zijn, zelfs al beschikte hij over onbeperkte middellen en tijd om tot deze overtuigingen te komen. De overtuiging van de expert is niet volledig bepaald in de zin dat hij niet op elke vraag een antwoord weet, of soms onbeslist is. Het klinkt dan ook redelijk om te zeggen dat de kennis van een expert, tot op zeker hoogte, inherent imprecies is [62, Hoofdstuk 5]. Deze simpele observatie doet niet alleen de sensitiveitsanalyse falen, maar toont ook aan dat gewone waarschijnlijkheden ontoereikend zijn om alle finesses van het modelleren van onzekerheid te bevatten: modellen gebaseerd op klassieke waarschijnlijkheden kunnen immers nooit onbeslist zijn.

Aangezien waarschijnlijkheden dus te kort schieten, stellen we in hoofdstuk 1_{28} een nieuw model voor ter vervanging van klassieke waarschijnlijkheden. Dit model wordt een **coherente verzameling van aanvaardbare gokken** genoemd en is gestoeld op de Finetti's operationele subjectivistische aanpak [24]. Een verzameling van aanvaardbare gokken wordt $\mathcal{A}$ genoteerd. We stellen een operationele en gedragsgerichte aanpak voorop omdat we willen dat het model een duidelijke interpretatie heeft en de consequenties van een beoordeling goed gedefinieerd zijn. Het operationele aspect van de modellering van de overtuigingen van een expert in verband met de mogelijke uitkomsten van een toevallige veranderlijke, bestaat er in om de expert een aantal gokken aan te bieden. Een gok f is een reëelwaardige functie gedefinieerd op een **eindige mogelijkhedenruimte** die overeenkomt met de gemodelleerde toevallige veranderlijke. De expert kan dan voor elke aangeboden gok f besluiten of hij deze aanvaardbaar vindt ($f \in \mathcal{A}$) of niet ($f \notin \mathcal{A}$). Als hij ze aanvaardbaar vindt, dan wil dit zeggen dat hij geen bezwaar ziet, om deze gok en de gevolgen ervan aan te nemen. Een gok aanvaarden is dus een beslissing nemen in onzekerheid.

Uiteraard wensen we dat de expert tegemoet komt aan een aantal rationaliteitseisen. Zo wordt er bijvoorbeeld gesteld dat het niet rationeel is, om een gok te aanvaarden die gegarandeerd verlies oplevert. Als aan de rationaliteitsaxioma's voldaan is, dan noemen we de verzameling van aanvaardbare gokken **coherent**. De gebruikte rationaliteitsaxioma's zijn:

A1. $\mathcal{A} \cap \mathcal{L}_{<0} = \emptyset$ [zeker verlies vermijden]

A2. $\mathcal{L}_{\geq 0} \subseteq \mathcal{A}$ [gedeeltelijke winst aannemen]

A3. $\text{posi}(\mathcal{A}) = \mathcal{A}$ [het deductieprincipe voor aanvaardbaarheid]

Uiteraard belet niets ons om gokken die afhangen van meerdere toevallige veranderlijken aan te bieden aan de expert en onze theorie is dus effectief in staat om multivariate onzekerheidsmodellen te beschrijven. We tonen ook aan hoe we kunnen marginaliseren en conditioneren en zo krijgt de theorie van de coherente verzamelingen van aanvaardbare gokken de allures van een voldragen onzekerheidsmodelleringstheorie. We kunnen ons ook afvragen hoeveel de expert bereid zou zijn te betalen voor een gok die hem aangeboden wordt. Deze maximale (supremale) aanvaardbare prijs wordt de **onderprevisie** $\underline{\mathrm{P}}(f)$ van de gok f genoemd en we tonen aan hoe deze onderprevisie afgeleid kan worden, wanneer een coherente verzameling van aanvaardbare gokken gegeven wordt:

$$\underline{\mathrm{P}}(f) = \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{A}\}.$$

Alternatief kunnen we ons afvragen welke de minimale (infimale) aanvaardbare prijs voor de expert is, opdat hij een gok zou verkopen. Deze minimale prijs wordt de **bovenprevisie** $\overline{\mathrm{P}}(\cdot)$ genoemd. Het verband tussen de onder- en bovenprevisie wordt gegeven door $\overline{\mathrm{P}}(f) = -\underline{\mathrm{P}}(-f)$.

Dit werk is niet het eerste dat gebruik maakt van verzamelingen van gokken om overtuigingen van een expert te beschrijven. Bijvoorbeeld Williams [67] en voornamelijk Walley [62] definieerden al tal van verzamelingen van gokken. Een van de redenen waarom we toch ons eigen model uitwerkten, is vanwege de stiefmoederlijke behandeling van de nulgok, de gok die niets oplevert, maar ook geen verlies met zich meebrengt. Ook de behandeling van onverschilligheid laat te wensen over. Zo definieerde Walley bijvoorbeeld twee verschillende soorten van coherente verzamelingen van wenselijke gokken, die zich van mekaar onderscheiden door de nulgok wel, volgens de eerste definitie, en niet, volgens de tweede definitie als wenselijk te beschouwen. Het probleem is het niet onderkennen van de fundamentele onverschilligheid die gepaard zou moeten gaan met de nulgok. Als de expert de nulgok aangeboden zou krijgen, dan zou hij geen reden mogen zien om hem niet te aanvaarden, maar als de expert de nulgok al had, dan zou hij evenmin een reden mogen zien om deze nulgok van de hand te doen. De **verzameling van onverschillige gokken** $\mathcal{I}$ is gedefinieerd als die gokken die aanvaardbaar zijn, terwijl ook hun negatie aanvaardbaar is: $\mathcal{I} = \mathcal{A} \cap -\mathcal{A}$. Een groot voordeel van het expliciet maken van deze categorie, is dat we een elegante manier hebben om (sterke) symmetrie te beschrijven. Als de expert bijvoorbeeld denkt dat de permutatie van de opeenvolging van de uitkomsten van een experiment er niet toe doet, dan impliceert dit dat hij onverschillig is wanneer hij moet kiezen tussen een gok f en zijn permutatie πf: $f - \pi f \in \mathcal{I}$. De behandeling van symmetrie, die evenwel niet centraal staat in deze thesis, is dan ook iets waar onze coherente verzamelingen van aanvaardbare gokken in uitblinken.

Een van de eenvoudigste grafische waarschijnlijkheidsmodellen zijn waarschijnlijkheids- of gebeurtenissenbomen en het is dan ook niet onlogisch om ons onderzoek naar grafische operationele onzekerheidsmodellen te starten bij deze structuur. In hoofdstuk 2_{80} veralgemenen we waarschijnlijkheidsbomen door de waarschijnlijkheden in de nodes te vervangen door coherente verzamelingen van aanvaardbare gokken, of door onderprevisies, en noemen het resultaat een **imprecieze gebeurtenissenboom** [12]. De nodes in dit bijzonder soort grafen zijn de situaties, of mogelijke uitkomsten, waarin het gemodelleerde systeem zich kan bevinden. De boomstructuur legt de opeenvolging (volgorde) van de situaties vast en kan gezien worden als het vastleggen van causale verbanden [53]. De coherente verzamelingen van accepteerbare gokken in de nodes zijn lokale modellen geassocieerd met knooppunten die de overtuiging van de expert over direct volgende situaties beschrijven. We tonen hoe deze lokale modellen samengesteld kunnen worden tot een globaal model en belangrijker, geven een algoritme dat de onderprevisie van een globale gok op een efficiënte manier kan bepalen. Hierbij wordt dan wel verondersteld dat de diepte van de boom eindig is.

Alhoewel het concept imprecieze gebeurtenissenboom vrij eenvoudig lijkt, mag deze ontwikkeling toch niet onderschat worden, daar zij een eerste stap vormt in het ontwikkelen van een gebeurtenisgedreven theorie van toevalsprocessen in discrete tijd, gebaseerd op imprecieze waarschijnlijkheden. Als voorbeeld behandelen we het gekende gokkers-bankroet (gambler's ruin) probleem, bespreken we een eenvoudige veralgemening van het Bernoulliproces en leiden we uitdrukkingen af voor veralgemeende identieke en onafhankelijk verdeelde processen (of steekproeven). Er bestaat een sterk verband [12] tussen onze imprecieze waarschijnlijkheidsbomen en Shafer en Vovk's speltheoretische waarschijnlijkheidstheorie [55]. In deze laatste theorie staan de concepten sub- en supermartingaal centraal en we definiëren deze speciale processen ook in ons raamwerk. Deze sub-en supermartingalen worden door Shafer en Vovk bijvoorbeeld uitvoerig gebruikt om limietwetten af te leiden en we gaan na of Shafer en Vovk's methodes overdraagbaar zijn naar imprecieze gebeurtenissenbomen. Als voorbeeld leiden we de zwakke wet van de grote getallen en de Hoeffding-Azuma-vergelijking af.

Een ander grafische model, dat in hoofdstuk 4_{160} van dit proefschrift beschreven wordt, is de **imprecieze Markovboom** [13]. Net zoals imprecieze gebeurtenissenbomen, is de grafische structuur een boom, maar in tegenstelling tot gebeurtenissenbomen waar de nodes situaties representeren, zijn de nodes in Markovbomen toevallige veranderlijken. De locale modellen worden nu niet gegeven in de vorm van coherente verzamelingen van aanvaardbare gokken, maar we beperken ons in dit hoofdstuk tot coherente onderprevisies, conditioneel op de voorgaande node (de ouder- of moederknoop). De vraag die zich nu stelt is: wat de betekenis is van deze Markovboom? Klassieke Bayesiaanse netwerken coderen d-scheiding. Van d-scheiding is echter geweten [10] dat ze zeker niet overdraagbaar is naar imprecieze waarschijnlijkheden omdat ze symmetrie codeert die in imprecies probabilistische modellen slechts uitzonderlijk gegarandeerd is. Daarom geven we de grafische structuur een licht andere betekenis: conditioneel op de ouders zijn alle toevallige veranderlijken die niet op de ouders volgen (de ouders niet meegerekend) **epistemisch irrelevant** [42, 62] voor de nodes die wel op de ouders volgen. In een boom is er natuurlijk telkens maar één ouder die we dan de moeder noemen. We zeggen dat een veranderlijke X epistemisch irrelevant is voor Y wanneer het observeren van de waarde van X, geen invloed heeft op de model dat de onzekerheid beschrijft voor Y. Dit is nieuw. In tegenstelling tot onze aanpak, veronderstelt het leeuwendeel van de imprecieze grafische modellen die tot nog toe bestudeerd werden sterke onafhankelijkheid in plaats van epistemische irrelevantie. Dit soort grafische modellen wordt ook wel credale netwerken genoemd, en is sterk gelinkt met sensitiviteitsanalyse. Het imprecieze model wordt dan beschouwd als de gedeeltelijke beschrijving van een uniek precies, klassiek waarschijnlijkheidsmodel.

Ook voor imprecieze Markovbomen zijn we in staat om een globale onderprevisie op te bouwen uit de lokale onderprevisies. Computationeel is deze uitdrukking echter niet onmiddellijk toepasbaar. We kunnen de geconstrueerde, globale onderprevisie echter wel gebruiken om op een efficiënte wijze de onderprevisie van een gok op één bepaalde node, conditioneel op een willekeurig aantal geobserveerde nodes te bepalen. Dit exacte algorithm gebruikt net als Pearls bekende algoritme [47] een techniek van message-passing, maar is toch fundamenteel en conceptueel verschillend.

Als laatste grafische model bespreken we in hoofdstuk 3$_{122}$ de **imprecieze Markovketen** [14, 28, 29, 59], die zowel kan geïnterpreteerd worden als een speciaal soort imprecieze gebeurtenissenboom, en als een speciaal type imprecieze Markovboom. Een imprecieze Markovketen is een imprecieze Markovboom, waarbij de grafische structuur—zoals de naam aangeeft—een ketting vormt. Het scheidingsprincipe, dat de grafische structuur van de imprecieze Markovboom betekenis gaf, wordt in dit geval de Markovvoorwaarde genoemd. Deze Markovvoorwaarde zegt dat alle knopen die voorafgaan aan een bepaalde knoop, epistemisch irrelevant zijn voor al de nodes die volgen op deze bepaalde knoop, en dit op voorwaarde dat de waarde van deze ene node gekend is. De Markovconditie rechtvaardigt het gebruik van het woord toestand in imprecieze Markovketens. Het blijkt bovendien voordelig om de lokale conditionele bovenwaarschijnlijkheden samen te vatten in één boventransitieoperator, het imprecieze analogon van de transitiematrix in klassieke Markovketens. Een imprecieze Markovketen wordt dan beschreven door zijn boventransitieoperator (of boventransitieoperatoren in het geval de Markovketen niet stationair is) en een initiële bovenprevisie. Bij de keuze van de Markovconditie hadden we ook weer sterke onafhankelijkheid in plaats van epistemische irrelevantie kunnen gebruiken. We tonen aan dat zowel epistemische irrelevantie als sterke onafhankelijkheid tot dezelfde onderprevisie op marginale gokken leiden, we geven een algoritme dat lineair is in het aantal beschouwde knopen.

Een belangrijk geval waarbij we enkel marginale gokken bestuderen, is bij de studie van het limietgedrag van Markovketens. We slagen er in om een veralgemeende Perron-Frobeniusstelling te bewijzen en demonstreren bovendien dat deze eigenschap equivalent is met ergodiciteit. Ergodiciteit impliceert ook dat er precies één invariante (initiële) bovenprevisie is. We ontwikkelen een efficiënt algoritme dat kan bepalen of een imprecieze Markovketen ergodisch is. In het speciale geval van imprecieze Markovketens met een tweedimensionale toestandsruimte beschrijven we bovendien het gedrag van de Markovketen in termen van de eigenwaarden en vectoren van de boventransitieoperator.

Summary

The last twenty years have witnessed a rapid growth of **graphical models** in the fields of artificial intelligence and statistics. These models combine graphs and probability theory to address complex multivariate problems in a variety of domains, such as medicine, finance, risk analysis, defence, and environment, to name just a few. Often, the parameters of the graphical model are not known precisely, and that is why one considers the set of all the graphical models that are consistent with the partial specification of the parameters. Common causes for the existence of partial knowledge are the cost of, and time constraints on, eliciting parameters, and disagreement amongst a group of experts consulted for that purpose. Non-ignorable missing data can be another reason, in case the parameters are inferred from a data set.

The sensitivity analysis interpretation of imprecise probability models, and hence strong independence, is not always applicable. A notable case arises when one wishes to model an expert's beliefs: it is then not always tenable that there should be some ideal probability that models these beliefs, and that it is only because of our limited resources that we cannot define it precisely. Rather, it seems more reasonable to concede that expert knowledge may be inherently imprecise to some extent [62, Chapter 5]. This simple observation not only shows that sensitivity analysis can fail, it also states that probabilities lack the expressiveness that is necessary to model imprecision and indecision adequately.

To overcome the shortcomings of classical probabilities, we propose in

Chapter 1[28] a new model as a replacement for them. The new models are called **coherent sets of acceptable gambles** and find their roots in de Finetti's operational subjectivist philosophy [24]. We denote a set of acceptable gambles by $\mathcal{A}$. We want our framework to be behaviouristic and operational because we want a clear interpretation of the model. The operational aspect exists in the offering of a number of gambles to the expert who is modelling the uncertainty. A gamble f is a real valued function, defined on a **finite possibility space** $\mathcal{X}$ that corresponds to any variable X that is modelled. The expert can decide whether he is willing to accept the offered gamble ($f \in \mathcal{A}$) or not ($f \notin \mathcal{A}$). If he considers a gamble to be acceptable, then this means that he does not object to accept the gamble and its consequences.

It is evident that we want the expert to meet some rationality criteria. It is assumed, for example, that accepting a gamble that guarantees sure loss is not a rational thing to do. If the rationality requirements are satisfied, then we say that the set of acceptable gambles is **coherent**. The rationality criteria we impose are:

A1. $\mathcal{A} \cap \mathcal{L}_{<0} = \emptyset$ [avoiding sure loss]

A2. $\mathcal{L}_{\geq 0} \subseteq \mathcal{A}$ [accepting partial gain]

A3. $\operatorname{posi}(\mathcal{A}) = \mathcal{A}$ [deduction principle for acceptability]

Of course, nothing says that we cannot offer the expert gambles that depend on a number of variables and our theory is effectively capable of describing multivariate uncertainty models. In that case, we show how our uncertainty models can be marginalised and updated which pushes our theory towards a full-fledged framework for uncertainty modelling. We can also ask ourselves how much the expert would be willing to pay for a gamble that is offered to him. This maximal (supremal) acceptable price is what we call the **lower prevision** $\underline{\mathrm{P}}(\cdot)$ and we show what is the relation between lower previsions and sets of acceptable gambles:

$$\underline{\mathrm{P}}(f) = \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{A}\}.$$

The minimal (infimum) acceptable price the expert is willing to give for a gamble is what we call the **upper prevision** $\overline{\mathrm{P}}(\cdot)$. The upper prevision can be found from the lower prevision through the conjugacy relation $\overline{\mathrm{P}}(f) = -\underline{\mathrm{P}}(-f)$.

Our work is not the first that uses sets of gambles to describe an expert's beliefs. Williams [67] and in a more prominent way Walley [62], define multiple types of sets of coherent gambles. The reason why we do not adopt one of their types of sets, is because of the way the zero gamble is treated. This zero gamble is the gamble that results neither gain nor profit, a status quo. Walley, for example, defines two different types of coherent sets of desirable gambles, that differ only in that in one definition, the zero gamble is assumed to be desirable, and in the other it is not desirable. We believe that, if the zero gamble is offered to an expert, he should have no objections to buying it, nor should he object to selling it. In other words, the expert should be indifferent towards the zero gamble. We define the category of **indifferent gambles** $\mathcal{I}$ as the gambles that are acceptable themselves, but also their negation: $\mathcal{I} = \mathcal{A} \cap -\mathcal{A}$. A strong advantage of making this category of gambles explicit, is that it provides an elegant way of describing (strong) symmetry. If, for example, an expert believes that it does not matter, in a sequence of experiments, whether outcomes are permuted or not, then he can express this by saying that the gamble f minus his permuted version πf is an indifferent gamble: $f - \pi f \in \mathcal{I}$. The treatment of symmetry, although not central in this work, is where our approach to coherent sets of acceptable gambles stands out.

Probability (or event) trees are amongst the simplest graphical models and it is not illogical to start our research into graphical uncertainty models with this structure. In Chapter 280, we generalise event trees by replacing the probabilities in each node with coherent sets of acceptable gambles, or with coherent lower previsions. The result is an **imprecise probability tree** [12] and the nodes in this special type of graphical model represent the situations, or possible outcomes that the modelled system can be in. The coherent sets of acceptable gambles in the nodes are local models that describe the expert's belief about situations that follow immediately. We show how to combine these local belief models into a global model and importantly, give an algorithm that can calculate lower previsions of global gambles efficiently. In order for the algorithm to work, we have to assume that the depth of the tree is finite.

The concept of an imprecise probability tree sure seems simple, but this development should not be underestimated, as it is a first step towards an event-driven account of random processes in discrete time with imprecise probability models. As an example, we treat the famous gambler's ruin problem and derive an expression for a generalised version of independent and identically distributed processes (or random samples). There is a strong connection [12] between the theory we have developed, and Shafer and Vovk's game-theoretic probability [55]. In the latter, the concept of sub- and supermartingales is central and we define these special processes in our version as well. Shafer and Vovk use these methods extensively to prove generalisations of strong and weak laws from classical probability theory and we investigate whether the methods they use can be transferred to our imprecise event tree framework. As an example, we derive the Weak Law of Large Numbers and the Hoeffding - Azuma inequality.

Another graphical model, which we study in Chapter 4_{160}, is the **imprecise Markov tree** [13]. Like imprecise event trees, the graphical structure is a tree, but in contradistinction with imprecise event trees, where the nodes represent situations, the nodes in imprecise Markov trees are random variables. The local uncertainty models are given as coherent sets of acceptable gambles, but we restrict ourselves to lower previsions, conditional on the previous node (the parent or mother node). The question about the exact meaning of the graphical structure presents itself. Classical Bayesian networks encode d-separation. In the case of imprecise probability trees, we know however [10] that they do not satisfy d-separation because the symmetry it encodes can be guaranteed only in exceptional cases. That is why we have redefined the interpretation of the graphical model: conditionally on the parents, all random variables strictly preceding the parents are **epistemically irrelevant** [42, 62] to the random variables following the parents. We say that X is epistemically irrelevant to Y if observing X has no influence on the model that describes our beliefs about Y. The majority of the imprecise graphical networks that have been studied so far assume strong independence instead of epistemic irrelevance. The resulting graphical models are called credal networks and are strongly linked with sensitivity analysis. The imprecise uncertainty model is in that case considered a partial description of a precise, classical probability.

Also for imprecise Markov trees, we are able to build a global lower prevision from the local lower previsions. Computationally, this expression is not immediately applicable. We can use it, however, to efficiently calculate the lower prevision for a gamble on one particular node, conditional on any number of observed nodes. This exact algorithm bears strong similarities with Pearl's message-passing algorithm [47].

As a last graphical model, we describe **imprecise Markov chains** [14, 28, 29, 59] in Chapter 3_{122}. These imprecise Markov chains can be interpreted as a special type of imprecise probability trees, as well as a special type of imprecise Markov tree. An imprecise Markov chain is an imprecise Markov tree, where the graphical structure—as hinted by its name—is a chain. The separation principle that allows for the proper interpretation of the Markov tree, is in this special case called the Markov Condition. This Markov Condition states that all nodes, preceding a particular node are epistemically irrelevant for all nodes that follow this node on condition that the value in the assumed node is known. It is exactly this Markov Condition that allows for the use of the word "state" in Markov chains, as the state summarises all the information about the past (the preceding nodes). It turns out to be advantageous to summarise all the local upper previsions in a node in a single upper transition operator, the imprecise counterpart of the classical Markov chain's transition matrix.

We could also have assumed strong independence in the Markov condition. We show that both independence concepts lead to the same posterior marginal upper previsions and we give an algorithm that can compute such upper previsions with a complexity that is linear in the number of nodes.

An important case where marginal gambles are studied is the study of the limit behaviour of Markov chains. We succeed in proving a generalised version of the Perron - Frobenius theorem and show that the conditions that lead to this theorem are exactly the conditions that make the chain ergodic. Ergodicity implies that there is exactly one irrelevant (initial) upper prevision. We develop an efficient algorithm that can decide upon ergodicity of upper transition operators. In the special case of imprecise Markov chains with a two dimensional state space, we show how the behaviour of the Markov chain can be described completely in terms of the eigenvalues and eigengambles of the upper transition operator.

Introduction

Global overview

Probabilistic models are intended to represent an agent's beliefs about the world he is operating in, and which describe and even determine the actions he will take in a diversity of situations. Probability theory provides a normative system for reasoning and decision making in the face of uncertainty. Bayesian, or precise, probability models have the property that they are completely decisive: a Bayesian agent always has an optimal choice when faced with a number of alternatives, whatever his state of information. While many may view this as an advantage, it is not always realistic. Imprecise probability models try to deal with this problem by explicitly allowing for indecision, while retaining the normative, coherentist stance of the Bayesian approach. In Chapter 1, we develop our own uncertainty model which we call **coherent sets of acceptable gambles**. We follow the school of thought of Walley [62, 64] who follows the tradition of Frank Ramsey [48], Bruno de Finetti [24] and Peter Williams [69] in trying to establish a rational model for a subject's beliefs in terms of her behaviour.

Imprecise probability models appear in a number of AI-related fields. For instance in **probabilistic logic**: it was already known to George Boole [4] that the result of probabilistic inferences may be a set of probabilities (an imprecise probability model), rather than a single probability. This is also important for dealing with missing or incomplete data, leading to so-called partial identification of probabilities, see for instance [22, 38]. There is also a growing literature on so-called **credal nets** [8, 9]: these are essentially Bayesian nets with imprecise conditional probabilities.

We are convinced that it is mainly the mathematical and computational complexity often associated with imprecise probability models that is keeping them from becoming a more widely used tool for modelling uncertainty. But we believe that the results reported here can help make inroads in reducing this complexity. In Chapter 2, we develop a theory of **imprecise probability trees**: probability trees where the transition from a node to its children is described by an imprecise probability model in Walley's sense. Our results provide the necessary apparatus for making inferences in such trees. And because probability trees are so closely related to random processes, this effectively brings us into a position to start developing a theory of (event-driven) random processes where the uncertainty can be described using imprecise probability models.

We are able to prove so-called Marginal Extension results (Theorems 30 and 112, Proposition 97), which lead to backwards recursion, and dynamic programming-like methods that allow for an exponential reduction in the computational complexity of making inferences in such imprecise probability trees. For (precise) probability trees, similar techniques were described in Shafer's book on causal reasoning [53]. They seem to go back to Christiaan Huygens, who drew the first probability tree, and showed how to reason with it, in his solution to Pascal and Fermat's Problem of Points.

A special type of imprecise probability tree are **imprecise Markov chains**. Early work on the more mathematical aspects of modelling "imprecision" in Markov chains was done by Hartfiel [28] and Kozine & Utkin [36]. The main difference between these approaches and ours, is that the Markov condition is based on **epistemic irrelevance** instead of strong independence. More recently, Škulj [59] has begun a formal study of the time evolution and limit behaviour of such systems. For the imprecise Markov chains we define, we give in Section 3.6 the necessary and sufficient conditions for a generalised Perron - Frobenius theorem and prove furthermore that these conditions make the imprecise Markov chain **ergodic**. Similar work, but coming from a different background, was already done by Akian and Gaubert [1]. Using the alternative characterisation of ergodicity developed in Section 3.6 however, we are able in Section 3.8 to avoid the possibly critical—in terms of computational complexity—step of Akian and Gaubert's algorithm in [1, Section 6.3]: the computation of the subdifferential, which relies heavily on extreme points. Our newly designed algorithm is linear in the dimension of the state space, where the evaluation of the transition map is considered as an oracle. In Section 3.9 we prove that ergodicity is equivalent to a contraction property in Hilbert's seminorm which is related to the one previously followed by Škulj and Hable [61]. We explain the advantages and disadvantages of characterisation of ergodicity in terms of a coefficient of ergodicity.

As a last type of graphical model, we focus on credal nets, which are graphical models that generalise Bayesian nets to imprecise probabilities. We replace the notion of strong independence normally used in credal nets with the weaker notion of epistemic irrelevance. Focusing on directed trees, we call the resulting graphical models **imprecise Markov trees**. We show how to combine the given local uncertainty models in the nodes of the graph into a global model, and we use this to construct and justify an exact message-passing algorithm that computes updated beliefs for a variable in the tree.

Basic nomenclature

For two real-valued functions f and g on a finite set $\mathcal{S}$ we say that $f \leq g$ if and only if $f(x) - g(x) \leq 0$ for all $x \in \mathcal{S}$. If in addition $f \neq g$ then we write $f < g$. We define the stronger relation $\prec$ by $f \prec g \Leftrightarrow \max(f - g) < 0$ and we say that f is pointwise strictly smaller than g. Furthermore $f \geq g \Leftrightarrow g \leq f$, $f > g \Leftrightarrow g < f$ and $f \succ g \Leftrightarrow g \prec f$.

In this work, beliefs about variables are described. A random variable X assumes values in a possibility space that throughout this thesis is assumed to be finite. Any subset of the possibility space is called an **event**. A **gamble** f about X is a real function on $\mathcal{X}$ and it represents uncertain rewards, i.e., $f(x)$ specifies the amount of utility one gets if the random variable X assumes the value $x \in \mathcal{X}$. The set of all possible gambles about X is denoted by $\mathcal{L}(X)$ and $\mathcal{L}_{\geq 0}(X)$ represents the set of non-negative gambles $f \geq 0$. The set $\mathcal{L}_{>0}(X)$ is equal to $\mathcal{L}_{\geq 0}(X) \setminus \{0\}$. The set $\mathcal{L}_{\succ 0}(X)$ is equal to the set of all gambles f about X that are pointwise strictly greater than zero. Derived sets are $\mathcal{L}_{\leq 0}(X) := -\mathcal{L}_{\geq 0}(X)$, $\mathcal{L}_{<0}(X) := -\mathcal{L}_{>0}(X)$ and $\mathcal{L}_{\prec 0}(X) := -\mathcal{L}_{\succ 0}(X)$.

A special type of gamble that will often be used is the **indicator** I_A of a set A: it returns one on A and zero elsewhere. When appropriate, we will denote the indicator of a singleton $\{x\}$ also as I_x instead of $I_{\{x\}}$.

The concept of sets of gambles is central in this work. Whenever a gamble $f \in \mathcal{L}(X)$, we will assume throughout that also αf will be in $\mathcal{L}(X)$ where $\alpha \in \mathbb{R}$.When given two sets of gambles $\mathcal{F}$ and $\mathcal{G}$, then we can take the Minkowski sum of both sets which will be denoted $\mathcal{F}+\mathcal{G}$. Thus in general we can write for any $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$ and $\mathcal{F},\mathcal{G} \subseteq \mathcal{L}(X)$

$$\alpha\mathcal{F}+\beta\mathcal{G} := \{\alpha f+\beta g: f \in \mathcal{F} \text{ and } g \in \mathcal{G}\}.$$

One equivalence is often used in proofs and gets special mention.

$$(\mathcal{A}+\mathcal{B}) \cap \mathcal{C} = \varnothing \Leftrightarrow \mathcal{A} \cap (\mathcal{C}-\mathcal{B}) = \varnothing. \tag{1}$$

The restriction of a gamble $f \in \mathcal{L}(\mathcal{X})$ to the domain $\mathcal{S} \subseteq \mathcal{X}$ is denoted by $f|_{\mathcal{S}}$ and results in a gamble in $\mathcal{L}(\mathcal{S})$:

$$f|_A(x) := f(x) \text{ if } x \in \mathcal{S}.$$

Let $f,g \in \mathcal{L}(\mathcal{X})$. Then we denote by $\min f$ the minimal value $f(x)$ for $x \in \mathcal{X}$. The result of the operation $\min\{f,g\}$, on the other side, is again a gamble $h \in \mathcal{L}(\mathcal{X})$ defined by $h(x) = \min\{f(x),g(x)\}$. This is an exception of the case $\min\mathcal{S}$ which in general stands for the minimal elements of the partial order $(\mathcal{S},\leq)$ (see also Appendix B).

CHAPTER 1

Modelling uncertainty

1.1 Introduction

This[1] chapter explains how uncertainty can be modelled using an extended version of the theory of imprecise probabilities [62], or what might be called prevision, or acceptability, theory. This theory follows the subjective betting interpretation of de Finetti [23] but rather than working with previsions, the theory adopts William's [67] acceptable bets idea. Often, a partial [32] or strict [51] preference ordering of bets is used as a basic notion in uncertainty modelling. This path is not followed here because an operational approach to uncertainty modelling is favoured. The operationalism manifests itself in terms of simple questions about rewards on possible outcomes of experiments—called gambles—asked to an assessor whose belief is modelled. Depending on the question asked, the sets of gambles get different names and with these different names, different rationality axioms come as well. Walley, for example, defines sets of desirable, almost desirable, really and strictly desirable gambles. One of the inconveniences with these different sets of gambles is the arbitrariness of the inclusion or exclusion of the zero gamble. Rather than having a clear behavioural interpretation, the inclusion depends mostly on the simplicity wanted from the updating rule. In this chapter the zero gamble is—possibly together with other gambles—granted the special status it deserves and by doing so, a new theory of uncertainty

[1] Footnote dedicated to Enrique Miranda:

modelling is formed that encompasses several previous existing theories as a special case.

1.2 Acceptability, Indifference and Desirability

1.2.1 Modelling through sets of gambles

How can beliefs be specified about the possible values of a variable X that may assume values in a set $\mathcal{X}$? In traditional probability theory this is done using probabilities, where usually the probability $\mathrm{P}(X = x)$ is given for every element x of $\mathcal{X}$. Another way of formalising beliefs about X in a behavioural way uses **sets of gambles**. This is what is set out in this section.

A gamble f on X can be thought of as a reward $f(x)$ obtained when X assumes the value x. If a subject—also called assessor—is offered a particular gamble, then he or she might consider whether to **accept** the gamble or not. Accepting or not will depend on whether the subject in question expects to gain utility from the gamble and therefore, it says something about the subject's beliefs about X. Of course there is the possibility that the assessor is undecided about whether to accept f or not. If time were not the issue, it would be theoretically possible to present all gambles to the assessor and ask about his opinion regarding acceptability. This divides the set of all gambles into three subsets: the **set of acceptable gambles** $\mathcal{A}$,[2] the **set of declined gambles** $\mathcal{U}$ and the **set of unresolved gambles** $\mathcal{O}$, and sets a first axiom:

$$\{\mathcal{A}, \mathcal{O}, \mathcal{U}\} \text{ partitions } \mathcal{L}(X). \tag{1.1}$$

The mere existence of the set of unresolved gambles shows one of the great advantages of the theory of acceptable gambles: it incorporates a natural framework for decision that leaves space for indecision due to, for example, a lack of knowledge.

Before going on, it is useful to consider the gamble $-f$ that pays off the negation of f meaning that if X assumes the value x, then the owner of the gamble will gain the value $-f(x)$. Gaining $-f(x)$ can also be interpreted as losing $f(x)$, which imposes some extra behavioural structure on the set of acceptable gambles. If a gamble f is acceptable then its negative can be considered acceptable or not. The former leads to the **set of indifferent**

[2] The modelled random variable X can be included as a subscript in the notation if it is unclear from the context which variable is modelled, leading to the notation $\mathcal{A}_X$.

gambles $\mathcal{I} = \mathcal{A} \cap -\mathcal{A}$ and the latter to the **set of desirable gambles** $\mathcal{D} = \mathcal{A} \setminus -\mathcal{A}$. If a gamble f is classified as indifferent then the assessor is indifferent to accepting or declining the gamble f. Strictly speaking, there could be a third possibility, which has been omitted here: if a gamble f is accepted then the assessor could choose not to say anything about $-f$. However, assuming linear utility, getting a gamble f has exactly the same value as giving away the gamble $-f$ which would make it odd for an assessor to have a distinct opinion about accepting a gamble, but not about declining it. In short, we assume that

$$\mathcal{U} = -\mathcal{D}. \tag{1.2}$$

In general, it is wise for the assessor to meet a minimal number of rationality requirements. For example, not accepting a strictly positive gamble would commonly be considered as irrational behaviour as this gamble gives a sure gain. We sublimate rationality for sets of desirable gambles $\mathcal{D}$ and sets of indifferent gambles $\mathcal{I}$ into a set of axioms. If a couple $(\mathcal{I}, \mathcal{D})$ meets these axioms, then we say that the couple is **coherent**. Assuming **linear utility**, the following axioms express the coherence requirements for a set of desirable gambles $\mathcal{D}$ and a set of indifferent gambles $\mathcal{I}$.

Definition 1: *If two sets $\mathcal{I}$ and $\mathcal{D}$ of gambles on a finite space $\mathcal{L}(X)$ satisfy:*

ID1. $\mathcal{D} \cap \mathcal{I} = \emptyset$ [resolvability]

ID2. $\mathcal{D} \cap \mathcal{L}_{\leq 0} = \emptyset$ [not desiring partial loss]

ID3. $\mathcal{L}_{>0} \subseteq \mathcal{D}$ [desiring sure gain]

ID4. $\mathcal{L}_{\geq 0} \subseteq \mathcal{D} \cup \mathcal{I}$ [accepting partial gain]

ID5. $\operatorname{posi}(\mathcal{D}) = \mathcal{D}$ [deduction principle for desirability]

ID6. $\operatorname{span}\mathcal{I} = \mathcal{I}$ [deduction principle for indifference]

ID7. $\mathcal{D} + \mathcal{I} \subseteq \mathcal{D}$ [desiring sweetened deals]

then we call $\mathcal{I}$ a set of indifferent gambles and $\mathcal{D}$ a set of desirable gambes and call the couple $(\mathcal{I}, \mathcal{D})$ coherent.

Here, the **positive hull** of a set $\mathcal{S}$, $\text{posi}(\mathcal{S})$, stands for the set of all possible positive linear combinations of elements of $\mathcal{S}$,

$$\text{posi}(\mathcal{S}) := \left\{ \sum_{f \in A} \alpha_f f \colon A \subseteq \mathcal{S}, |A| \in \mathbb{N}_{>0} \text{ and } \alpha_f \in \mathbb{R}_{>0} \right\}, \tag{1.3}$$

and the **linear span** $\text{span}\,\mathcal{S}$ is the set of all linear combinations of elements of $\mathcal{S}$,

$$\text{span}\,\mathcal{S} := \left\{ \sum_{f \in A} \alpha_f f \colon A \subseteq \mathcal{S}, |A| \in \mathbb{N}_{\geq 0} \text{ and } \alpha_f \in \mathbb{R} \right\}. \tag{1.4}$$

$|A|$ stands for the cardinality of the set A. By using $\mathcal{L}_{\leq 0}$ instead of $\mathcal{L}_{<0}$ in Axiom ID2 the situation is avoided where partial loss is accepted and partial gain declined. A consequence of these axioms is that there should be no indifference to sure loss:

$$\mathcal{I} \cap \mathcal{L}_{<0} = \varnothing. \tag{1.5}$$

This is so because an indifferent gamble in $\mathcal{L}_{<0}$ would require an indifferent gamble in $\mathcal{L}_{>0}$ which collides with axiom ID3. Axiom ID6 implies that the zero gamble should be indifferent. Together with Axiom ID5, this tells us that the set of desirable gambles $\mathcal{D}$ constitutes a convex cone that does not contain zero. A consequence of coherence is that it is not possible that both f and $-f$ are desirable. It is not too difficult to see that $\mathcal{D} \cup \mathcal{I}$ constitutes a convex cone that contains zero.

Axiom ID7 is important because it homes in on the true meaning of indifference. It states that the combination of an indifferent gamble with a gamble of a particular category will inherit this category.

Lemma 2 (interpretation of indifference): *Consider a couple of coherent sets of indifferent and desirable gambles* $(\mathcal{I}, \mathcal{D})$*. For any* $\mathcal{S} \subseteq \mathcal{L}(X)$ *with* $\mathcal{I} + \mathcal{S} \subseteq \mathcal{B}$ *where* $\mathcal{B}$ *is any element of the set* $\{\mathcal{D}, \mathcal{U}, \mathcal{I}, \mathcal{O}, \mathcal{A}\}$*, it holds that* $\mathcal{S} \subseteq \mathcal{B}$*.*

Proof: We know from ID7 that $\mathcal{D} + \mathcal{I} \subseteq \mathcal{D}$ and because $0 \in \mathcal{I}$ by ID6 we also know that $\mathcal{D} + \mathcal{I} \supseteq \mathcal{D}$ whence

$$\mathcal{D} + \mathcal{I} = \mathcal{D}. \tag{1.6}$$

If follows immediately from Equation (1.2) and ID6 that also $\mathcal{U} + \mathcal{I} = \mathcal{U}$. We did already know by ID6 that $\mathcal{I} + \mathcal{I} = \mathcal{I}$ and because $\mathcal{A} = \mathcal{I} \cup \mathcal{D}$ it also holds that $\mathcal{A} + \mathcal{I} = \mathcal{A}$. Finally, using the partitioning property 1.1[29] we get that $\mathcal{O} + \mathcal{I} = \mathcal{O}$. Using these findings together with $\mathcal{S} + \mathcal{I} \subseteq \mathcal{B}$ and $0 \in \mathcal{I}$ it follows immediately that $\mathcal{S} \subseteq \mathcal{B}$. □

A complete belief specification about a random variable X involves classifying all gambles in $\mathcal{L}(X)$ into acceptable $\mathcal{A}$, indifferent $\mathcal{I}$ or undesirable gambles $\mathcal{U}$. Because we assume that $\mathcal{U} = -\mathcal{D}$ by (1.2) and because by Equation (1.1) the sets of desirable, indifferent, undesirable and unresolved gambles are disjoint, it is sufficient to specify the sets $\mathcal{D}$ and $\mathcal{I}$ to model the subject's beliefs. Even stronger, given a set of acceptable gambles $\mathcal{A}$, we are able to recover $\mathcal{D}$, $\mathcal{U}$, $\mathcal{I}$ and $\mathcal{O}$.

Proposition 3: *A set of acceptable gambles $\mathcal{A}$ or a couple $(\mathcal{I},\mathcal{D})$ fixes the sets of acceptable $\mathcal{A}$, indifferent $\mathcal{I}$, desirable $\mathcal{D}$, undesirable $\mathcal{U}$ and unresolved $\mathcal{O}$ gambles in a unique way.*

Given $\mathcal{A}$ we have that

$$\begin{aligned}
\mathcal{D} &= \mathcal{A} \setminus -\mathcal{A}, \\
\mathcal{I} &= \mathcal{A} \cap -\mathcal{A}, \\
\mathcal{U} &= -\mathcal{D} = -\mathcal{A} \setminus \mathcal{A}, \\
\mathcal{O} &= (\mathcal{A} \cup -\mathcal{A})^c = \mathcal{L}(X) \setminus (\mathcal{A} \cup -\mathcal{A}).
\end{aligned}$$

Given $(\mathcal{I},\mathcal{D})$ we have that

$$\begin{aligned}
\mathcal{A} &= \mathcal{D} \cup \mathcal{I}, \\
\mathcal{U} &= -\mathcal{D}, \\
\mathcal{O} &= (\mathcal{D} \cup \mathcal{I} \cup -\mathcal{D})^c = \mathcal{L}(X) \setminus (\mathcal{D} \cup \mathcal{I} \cup -\mathcal{D})
\end{aligned}$$

As a consequence, if given a coherent couple $(\mathcal{I},\mathcal{D})$ of sets of indifferent and desirable gambles, the corresponding set of acceptable gambles $\mathcal{A}$ is defined by $\mathcal{A} := \mathcal{I} \cup \mathcal{D}$. Conversely, if given a coherent set of acceptable gambles, the corresponding sets of indifferent and acceptable gambles are defined as $\mathcal{I} := \mathcal{A} \cap -\mathcal{A}$ and $\mathcal{D} := \mathcal{A} \setminus -\mathcal{A}$. So there is a one to one relation between sets of acceptable gambles and couples $(\mathcal{I},\mathcal{D})$. This means that it is possible to rewrite the rationality axioms of Definition 1_{30} in terms of $\mathcal{A}$ alone.

Definition 4: *A set of acceptable gambles $\mathcal{A}$ on a space $\mathcal{L}(X)$ is coherent if and only if*

A1. $\mathcal{A} \cap \mathcal{L}_{<0} = \emptyset$ [avoiding sure loss]

A2. $\mathcal{L}_{\geq 0} \subseteq \mathcal{A}$ [accepting partial gain]

A3. $\operatorname{posi}(\mathcal{A}) = \mathcal{A}$ [deduction principle for acceptability]

Essentially, both $\mathcal{A}$ and $(\mathcal{I},\mathcal{D})$ model exactly the same, so there should be only one way to match a coherent set $\mathcal{A}$ with a coherent couple $(\mathcal{I},\mathcal{D})$ and vice versa.

Proposition 5: *A set of acceptable gambles $\mathcal{A}$ is coherent if and only if the corresponding couple of indifferent and desirable gambles $(\mathcal{I},\mathcal{D})$ is coherent.*

Proof: We start by showing that a coherent set of acceptable gambles leads to a coherent couple $(\mathcal{I},\mathcal{D})$.

ID1. From $\mathcal{D} := \mathcal{A} \setminus -\mathcal{A}$ and $\mathcal{I} := \mathcal{A} \cap -\mathcal{A}$ it follows immediately that $\mathcal{D}$ and $\mathcal{I}$ are disjoint.

ID2. If $f \in \mathcal{L}_{\leq 0}$ then $-f \in \mathcal{A}$ by A2. Hence $f \in -\mathcal{A}$ so $f \notin \mathcal{D}$.

ID3. From A2 we get that $\mathcal{L}_{>0} \subseteq \mathcal{D} \cup \mathcal{I}$. If $f \in \mathcal{L}_{>0}$ then $-f \notin \mathcal{A}$ by A1. Hence $f \notin \mathcal{I}$ and therefore $\mathcal{L}_{>0} \subseteq \mathcal{D}$.

ID4. This follows immediately from A2.

ID5. It is sufficient to show that i) $\alpha(\mathcal{A} \setminus -\mathcal{A}) = (\mathcal{A} \setminus -\mathcal{A})$ for $\alpha \in \mathbb{R}_{>0}$ and, ii) for every $f, g \in (\mathcal{A} \setminus -\mathcal{A})$ it holds that $f + g \in \mathcal{A} \setminus -\mathcal{A}$.

i) From A3 it follows that $\alpha\mathcal{A} = \mathcal{A}$ and therefore also $-\alpha\mathcal{A} = -\mathcal{A}$. But this implies that $\alpha(\mathcal{A} \setminus -\mathcal{A}) = \mathcal{A} \setminus -\mathcal{A}$.

ii) Consider f and g in $\mathcal{D}$. Then both f and g belong to $\mathcal{A}$, and therefore $f + g \in \mathcal{A}$ by A3. Suppose *ex absurdo* that $-(f+g) \in \mathcal{A}$, then $-f = -(f+g) + g \in \mathcal{A}$, again by A3, which contradicts $f \in \mathcal{D}$.

ID6. It is again sufficient to show that i) $\alpha\mathcal{I} = \mathcal{I}$ for $\alpha \in \mathbb{R}$, and ii) for every $f, g \in \mathcal{I}$ it holds that $f + g \in \mathcal{I}$.

i) Consider $f \in \mathcal{I}$ and $\alpha \in \mathbb{R}$. There are three possibilities.

a) If $\alpha = 0$ then we see from A2 that $0 = \alpha f \in \mathcal{A}$ and because $-0 = 0$ also $\alpha f \in \mathcal{I}$.

b) If $\alpha > 0$ then $f \in \mathcal{A}$ implies $\alpha f \in \mathcal{A}$ and similarly $-f \in \mathcal{A}$ implies $-\alpha f \in \mathcal{A}$, by A3.

c) If $\alpha < 0$ then $f \in \mathcal{A}$ implies $-\alpha f \in \mathcal{A}$ and similarly $-f \in \mathcal{A}$ implies $\alpha f \in \mathcal{A}$, by A3.

ii) Consider f and g in $\mathcal{I}$. Then on the one hand f and g both belong to $\mathcal{A}$, and therefore $f + g \in \mathcal{A}$, by A3. And on the other hand $-f$ and $-g$ both belong to $\mathcal{A}$, and therefore $-(f+g) \in \mathcal{A}$, again by A3. Hence indeed $f + g \in \mathcal{I}$.

ID7. Consider $f \in \mathcal{D}$ and $g \in \mathcal{I}$. Then both f and g belong to $\mathcal{A}$, and therefore $f + g \in \mathcal{A}$ by A3. Suppose *ex absurdo* that $-(f+g) \in \mathcal{A}$, then $-f = -(f+g) + g \in \mathcal{A}$, again by A3, which contradicts $f \in \mathcal{D}$.

Next, we turn to the converse statement.

A1. Observe that

$$\mathcal{L}_{<0} \cap \mathcal{A} = \mathcal{L}_{<0} \cap (\mathcal{I} \cup \mathcal{D}) = \underbrace{(\mathcal{L}_{<0} \cap \mathcal{I})}_{\emptyset} \cup \underbrace{(\mathcal{L}_{<0} \cap \mathcal{D})}_{\emptyset \text{ by ID2}} = \emptyset.$$

To show that $\mathcal{L}_{<0} \cap \mathcal{I} = \emptyset$, combine ID130and ID330 into $\mathcal{L}_{>0} \cap \mathcal{I} = \emptyset$, and then apply ID630.

A2. $\mathcal{D} \cup \mathcal{I} = \mathcal{A}$ and ID430 prove this.

A3. $\text{posi}(\mathcal{A}) = \text{posi}(\mathcal{D} \cup \mathcal{I}) = \text{posi}(\mathcal{D}) \cup \text{posi}(\mathcal{I}) \cup \text{posi}(\mathcal{D} + \mathcal{I})$. From ID530, ID630 and ID730 we get that this is equal to $\mathcal{D} \cup \mathcal{I} \cup \mathcal{D}$, whence $\text{posi}(\mathcal{A}) = \mathcal{A}$. □

1.2.2 The consequences of an assessment

Although it is not feasible in practice to ask for the complete set of acceptable gambles $\mathcal{A}$, or alternatively, the complete sets of desirable and indifferent gambles, we can still try to deduce which coherent models are in accordance with—include—an **assessment** $\mathcal{A}_{as}$: a partial specification of $\mathcal{A}$. In an inference context, it is then interesting to know which of these coherent models—if any—is the least committal, where a set $\mathcal{A}_1$ of acceptable gambles is said to be **at most as committal** as the acceptable set $\mathcal{A}_2$ if $\mathcal{A}_1 \subseteq \mathcal{A}_2$. The least committal coherent model including the assessment will be called—if it exists—its natural extension.

Let us denote by $\mathbb{A}_X$ the **set of all coherent sets of acceptable gambles** on $\mathcal{L}(X)$. If there is no confusion, then we will write $\mathbb{A}$ instead of $\mathbb{A}_X$. If we provide this set with the natural partial ordering of set inclusion, we see that it has a smallest element, namely the set of all non-negative gambles $\mathcal{L}_{\geq 0}(X)$. It is not difficult to see that it is also closed under arbitrary non-empty intersections: $(\mathbb{A}, \subseteq)$ is a inf-semilattice where $\bigcap$ fulfils the role of infimum. From the coherence axioms of sets of acceptable gambles, we see that no gamble in the assessment $\mathcal{A}_{as}$ should be part of $\mathcal{L}_{<0}$.

Definition 6 (Avoiding sure loss): *Consider a subset $\mathcal{A}_{as}$ of $\mathcal{L}$. We say that this set $\mathcal{A}_{as}$ **avoids sure loss** if no positive linear combination of gambles in $\mathcal{A}_{as}$ is point-wise strictly negative:*

$$\text{posi}(\mathcal{A}_{as}) \cap \mathcal{L}_{<0} = \emptyset. \tag{1.7}$$

An assessment that does not avoid sure loss is said to incur sure loss. The next theorem shows that avoiding sure loss is the necessary and sufficient condition for an assessment to be extendable to a coherent model.

Theorem 7 (Natural extension): *Consider a set of acceptable gambles $\mathcal{A}_{as}$ on a space $\mathcal{L}$, and define its **natural extension**:*[3]

$$\operatorname{ext}(\mathcal{A}_{as}) := \bigcap\{\mathcal{A} \in \mathbb{A} : \mathcal{A}_{as} \subseteq \mathcal{A}\}. \tag{1.8}$$

Then the following statements are equivalent:

(i) $\mathcal{A}_{as}$ *avoids sure loss;*

(ii) $\mathcal{A}_{as}$ *is included in some coherent set of acceptable gambles;*

(iii) $\operatorname{ext}(\mathcal{A}_{as}) \neq \mathcal{L}$*;*

(iv) *the set of acceptable gambles* $\operatorname{ext}(\mathcal{A}_{as})$ *is coherent;*

(v) $\operatorname{ext}(\mathcal{A}_{as})$ *is the smallest coherent set of acceptable gambles that includes* $\mathcal{A}_{as}$.

When any (and hence all) of these equivalent statements hold, then

$$\operatorname{ext}(\mathcal{A}_{as}) = \operatorname{posi}(\mathcal{L}_{\geq 0} \cup \mathcal{A}_{as}) = \mathcal{L}_{\geq 0} + \operatorname{posi}(\mathcal{A}_{as} \cup \{0\}). \tag{1.9}$$

Proof: It follows from the fact that $\mathbb{A}$ is closed under arbitrary non-empty intersections, the definition of $\operatorname{ext}\mathcal{A}_{as}$, and the fact that $\mathcal{L}$ is not coherent, that the last four statements (ii)-(v) are equivalent.

Next, we prove that (i)$\Leftrightarrow$(ii).

(i)$\Leftarrow$(ii). Assume that $\mathcal{A}_{as}$ is included in some coherent set of acceptable gambles $\mathcal{A}$. Since $\mathcal{A} = \operatorname{posi}(\mathcal{A})$, $\mathcal{A}$ avoids sure loss by A1$_{32}$, and therefore so do all its subsets, including $\mathcal{A}_{as}$.

(i)$\Rightarrow$(ii) Conversely, assume that $\mathcal{A}_{as}$ avoids sure loss. For notational convenience, let $\mathcal{A}^* := \operatorname{posi}(\mathcal{L}_{\geq 0} \cup \mathcal{A}_{as})$. It is clear that $\mathcal{A}^*$ satisfies A2$_{32}$ and A3$_{32}$. Consider any $f \in \mathcal{A}^*$, so there are $n \in \mathbb{N}_{\geq 0}$, real $\lambda_k > 0$, $f_k \in \mathcal{A}_{as}$ and $g \geq 0$ such that $f = g + \sum_{k=1}^n \lambda_k f_k$. It follows from the assumption that $f - g \nless 0$ and therefore *a fortiori* $f \nless 0$, so $\mathcal{A}^*$ also satisfies A1$_{32}$, and is therefore coherent.

Finally, we prove that $\operatorname{ext}\mathcal{A}_{as} = \mathcal{A}^*$ whenever any (and hence all) of the equivalent statements (i)–(v) hold. Any coherent set of acceptable gambles that includes $\mathcal{A}_{as}$, must also include $\mathcal{A}^*$, by the axioms A2$_{32}$ and A3$_{32}$. Since we have proved above

[3] As commonly done, in this expression, we let $\bigcap\emptyset = \mathcal{L}$.

that $\mathcal{A}^*$ also satisfies A1$_{32}$ and is therefore coherent, it is the smallest coherent set of acceptable gambles that includes $\mathcal{A}_{as}$. Hence it is equal to $\text{ext}(\mathcal{A}_{as})$, by (v). The proof of the second equality in Equation (1.9) is trivial. □

In a more general form, an assessment can consist of sets of indifferent, desirable and acceptable gambles: $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$. In the quest for coherent sets of acceptable gambles $\mathcal{A}$ compatible with the assessment, it is not enough for the set $\mathcal{A}$ to dominate the assessment $\mathcal{A}_{as}$. Explicit mention of the indifferent and especially desirable gambles has to be taken into account.

Definition 8: *Given an assessment $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$, then a set of acceptable gambles $\mathcal{A}$ is said to **respect** the assessment if*

$$\begin{aligned} \mathscr{I}_{as} &\subseteq \mathcal{A} \cap -\mathcal{A}, \\ \mathcal{D}_{as} &\subseteq \mathcal{A} \setminus -\mathcal{A}, \\ \mathcal{A}_{as} &\subseteq \mathcal{A}. \end{aligned}$$

*We define the **associated set** of acceptable gambles $\tilde{\mathcal{A}}_{as}$ of an assessment $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ as*

$$\tilde{\mathcal{A}}_{as} := \mathscr{I}_{as} \cup -\mathscr{I}_{as} \cup \mathcal{D}_{as} \cup \mathcal{A}_{as}$$

*and the assessment is said to **avoid sure loss** if its associated set does:*

$$(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}) \text{ avoids sure loss } \Leftrightarrow \text{posi}(\tilde{\mathcal{A}}_{as}) \cap \mathcal{L}_{<0} = \emptyset.$$

Proposition 9: *A set of acceptable gambles $\mathcal{A}$ respects the assessment $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ if and only if $\tilde{\mathcal{A}}_{as} \subseteq \mathcal{A}$ and $\mathcal{A} \cap -\mathcal{D}_{as} = \emptyset$.*

Proof: If $\tilde{\mathcal{A}}_{as} = \mathscr{I}_{as} \cup -\mathscr{I}_{as} \cup \mathcal{D}_{as} \cup \mathcal{A}_{as} \subseteq \mathcal{A}$ and $\mathcal{A} \cap -\mathcal{D}_{as} = \emptyset$ then i) $\mathscr{I}_{as} \cup -\mathscr{I}_{as} \in \mathcal{A}$ whence $\mathscr{I}_{as} \in \mathcal{A} \cap -\mathcal{A}$, ii) $\mathcal{D}_{as} \in \mathcal{A}$ and $\mathcal{A} \cap -\mathcal{D}_{as} = \emptyset$ whence $\mathcal{D}_{as} \subseteq \mathcal{A} \setminus -\mathcal{A}$, and iii) $\mathcal{A}_{as} \in \mathcal{A}$, and we infer that the assessment is respected by $\mathcal{A}$. If on the other hand $\mathcal{A}$ respects the assessment, then it follows that

$$\tilde{\mathcal{A}}_{as} := \mathscr{I}_{as} \cup -\mathscr{I}_{as} \cup \mathcal{D}_{as} \cup \mathcal{A}_{as} \subseteq (\mathcal{A} \cap -\mathcal{A}) \cup (-\mathcal{A} \cap \mathcal{A}) \cup (\mathcal{A} \setminus -\mathcal{A}) \cup \mathcal{A} \subseteq \mathcal{A},$$

and $\mathcal{D}_{as} \cap -\mathcal{A} = \emptyset$ whence also $\mathcal{A} \cap -\mathcal{D}_{as} = \emptyset$. □

The following lemma shows that the set of acceptable gambles that respects an assessment constitutes a complete $\cap$-semilattice.

Lemma 10: *If $\mathcal{A}_0$ respects the assessment $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ and $\tilde{\mathcal{A}}_{as} \subseteq \mathcal{A}_i$ for all $i \in I$ where I is any index set, then $\mathcal{A}_0 \cap \bigcap_{i \in I} \mathcal{A}_i$ respects the assessment.*

Proof: If $\mathcal{A}_0$ respects the assessment then $\tilde{\mathcal{A}}_{as} \subseteq \mathcal{A}_0$ thus $\tilde{\mathcal{A}}_{as} \subseteq (\mathcal{A}_0 \cap \bigcap_{i \in I} \mathcal{A}_i)$. Moreover, if $-\mathcal{D}_{as} \cap \mathcal{A}_0 = \emptyset$ then surely $-\mathcal{D}_{as} \cap (\mathcal{A}_0 \cap \bigcap_{i \in I} \mathcal{A}_i) = \emptyset$ and the lemma follows from Proposition 9. □

The calculation of the natural extension of this more general type of assessment gets slightly more involved as the following theorem shows. Using Proposition 9 and Lemma 10, its proof is completely analogous to the proof of Theorem 7_{35}.

Theorem 11 (Natural extension): *Consider an assessment $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ and define its* ***natural extension*** *as*

$$\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}) := \bigcap \left\{ \mathcal{A} \in \mathbb{A} : \tilde{\mathcal{A}}_{as} \subseteq \mathcal{A} \text{ and } -\mathcal{D}_{as} \cap \mathcal{A} = \emptyset \right\}. \tag{1.10}$$

Then the following statements are equivalent.

(i) *there is some coherent set of acceptable gambles $\mathcal{A}$ that respects the assessment $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$;*

(ii) $\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ *is the smallest coherent set that respects the assessment $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$;*

(iii) $\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ *is coherent and $-\mathcal{D}_{as} \cap \operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}) = \emptyset$;*

(iv) $\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}) \neq \mathcal{L}$. *(the assessment cannot be respected)*

When any (and hence all) of these equivalent statements hold, then

$$\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}) = \mathcal{L}_{\geq 0} + \operatorname{posi}(\mathcal{D}_{as} \cup \mathcal{A}_{as} \cup \{0\}) + \operatorname{span} \mathscr{I}_{as}. \tag{1.11}$$

If a set of unresolved gambles $\mathcal{O}_{as}$ is given, then it could a posteriori be checked that $\operatorname{ext}(\mathcal{A}_{as}) \cap \mathcal{O}_{as} = \emptyset$. However, we would not mind too much if this test would fail; it would only imply that the assessor actually knows a bit more than he thought he knew.

▷ **Example 12:** One important example is the absence of any knowledge or information about a variable X in which case the set $\mathcal{A}_{as}$ is empty. This kind of model is called **vacuous** and $\operatorname{ext}(\mathcal{A}_{as}) = \mathcal{L}_{\geq 0}$, or equivalently, $(\mathcal{I}, \mathcal{D})$ is equal to $(\{0\}, \mathcal{L}_{>0})$.

▷ **Example 13:** Consider the random variable X with corresponding set of acceptable gambles

$$\mathcal{A} = \left\{ f \in \mathcal{L}(X) \colon \sum_{x \in \mathcal{X}} f(x) f_0(x) \geq \|f_0\|_2 \|f\|_2 \cos(\alpha_0) \right\}$$

with the gamble $f_0 \in \mathcal{L}_{>0}(X)$ and the angle α_0 given. Then this set of acceptable gambles will be coherent if and only if $\cos(\alpha_0)$ is chosen such that

$$\frac{f_0}{\|f_0\|_2} \geq \cos(\alpha_0) \geq 0.$$

This model could be interpreted as a neighbourhood model—typically used in sensitivity analysis—centred around f_0 and where $\cos(\alpha_0)$ measures the imprecision (inversely). The peculiar thing about this model is that it constitutes a convex cone of gambles with an infinite number of extreme rays.

1.2.3 Resolved models

Instead of asking which is the least-committal set extending an assessment $\mathcal{A}_{as}$, we could think about those sets that are compatible with the assessment and are most-committal. Given the definition of being "at most as committal as" in the previous section, these models coincide with the maximal elements (See Sections B.1194 and B.2196) of the partial order $(\mathbb{A}_{\supseteq \mathcal{A}_{as}}, \subseteq)$ where $\mathbb{A}_{\supseteq \mathcal{A}_{as}}$ is the set of all coherent sets of acceptable gambles that include $\mathcal{A}_{as}$:

$$\mathbb{A}_{\supseteq \mathcal{A}_{as}} \coloneqq \{\mathcal{S} \in \mathbb{A} \colon \mathcal{A}_{as} \subseteq \mathcal{S}\}.$$

This approach is perfectly fine when dealing with assessments in the form of a set of acceptable gambles. But when the assessment contains gambles explicitly labelled as desirable, then there can be undominated models that are not maximal elements of the partial order $(\mathbb{A}, \subseteq)$. Take for example the assessment depicted in Figure 1.1, where one gamble is considered acceptable, one gamble desirable and one gamble undesirable. The shaded region on the left is the natural extension, the shaded figure on the right is a resolved model, compatible with the assessment, that is not dominated by any maximal element of the partial order $(\mathbb{A}, \subseteq)$.
This is why we do not follow this approach here and rather call any set with no unresolved gambles ($\mathcal{O} = \emptyset$) maximally committal.

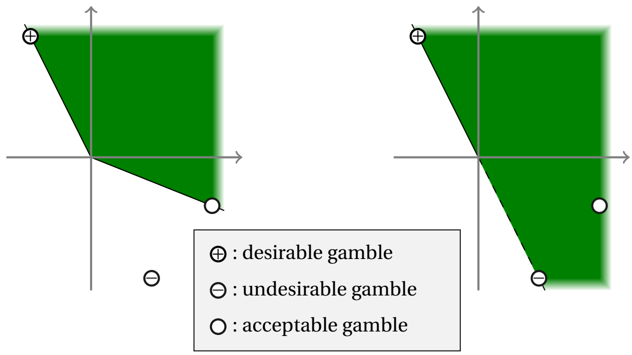


Figure 1.1: Example of the natural extension (left) and of a resolved model (right) that is not dominated by a maximal element of $(\mathbb{A}, \subseteq)$ compatible with the original assessment. The shaded region indicates the gambles that are considered acceptable in the model. Unlike a solid line, a dashed line on the border of the cone of acceptable gambles indicates that gambles on this line are not to be considered acceptable.

Definition 14: *For two sets of acceptable gambles $\mathcal{A}_1, \mathcal{A}_2 \in \mathcal{L}$ we say that $\mathcal{A}_1$ is **less resolved than** $\mathcal{A}_2$ and denote this by $\mathcal{A}_1 \sqsubseteq \mathcal{A}_2$ if*

$$\mathcal{A}_1 \sqsubseteq \mathcal{A}_2 \Leftrightarrow \mathcal{A}_1 \cup -\mathcal{A}_1 \subseteq \mathcal{A}_2 \cup -\mathcal{A}_2 \Leftrightarrow \mathcal{O}_1 \supseteq \mathcal{O}_2.$$

The relation "is less resolved than" on $\mathbb{A}$ is reflexive and transitive but not antisymmetric. Thus, $(\mathbb{A}, \sqsubseteq)$ is a partial preorder and the undominated elements of this preorder are what we call **resolved models**. They satisfy, and are characterised by, $\mathcal{O} = \emptyset$. Remark that $(\mathbb{A}, \sqsubseteq)$ is not a partial order as was $(\mathbb{A}, \subseteq)$. The next proposition gives a way of characterising resolved models.

Proposition 15: *The set of resolved models $\mathbb{M}$ can be characterised by*

$$\begin{aligned} \mathcal{M} \in \mathbb{M} &\Leftrightarrow \mathcal{M} \in \mathbb{A} \text{ and } \mathcal{M} \cup -\mathcal{M} = \mathcal{L} \\ &\Leftrightarrow \mathcal{M} \in \mathbb{A} \text{ and } (\forall f \in \mathcal{L})(f \notin \mathcal{M} \Rightarrow -f \in \mathcal{M}). \end{aligned}$$

The set of resolved models respecting an assessment $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ is denoted by $\mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})}$ and given by

$$\{\mathcal{M} \in \mathbb{M} : \mathcal{A}_{as} \subseteq \mathcal{M} \text{ and } \mathcal{D}_{as} \subseteq \mathcal{M} \setminus -\mathcal{M} \text{ and } \mathscr{I}_{as} \subseteq \mathcal{M} \cap -\mathcal{M}\}. \qquad (1.12)$$

We now intend to show that there is always at least one resolved model that includes a given coherent set of acceptable gambles. The following lemmas will be useful for doing so. The approach closely follows the route taken by Couso and Moral [6].

Lemma 16: *Consider an assessment* $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ *that avoids sure loss and an unresolved coherent set of acceptable gambles* $\mathcal{A}_0$ *that respects it. Then the set of acceptable gambles*

$$\mathcal{A}_1 \coloneqq \mathcal{A}_0 + \operatorname{span}\{f\}, \tag{1.13}$$

where f *is an unresolved gamble* ($f \notin \mathcal{A}_0 \cup -\mathcal{A}_0$) *is coherent and respects the assessment* $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$.

Proof: If f is unresolved then $f \notin \mathcal{A}_0 \cup -\mathcal{A}_0$ which means that the gamble $-f$ is also unresolved. Using Axioms ID6$_{30}$ and ID5$_{30}$ we infer that the set $\operatorname{span}\{f\} \setminus \{0\}$ is unresolved. If $\mathcal{A}_0 \cup (\operatorname{span}\{f\} \setminus \{0\})$ avoids sure loss, then we infer from Theorem 7$_{35}$ and the coherence of $(\mathcal{I}_0, \mathcal{D}_0)$ that the set of acceptable gambles $\mathcal{A}_1$ defined as

$$\begin{aligned}\mathcal{A}_1 &\coloneqq \operatorname{ext}(\mathcal{A}_0 \cup (\operatorname{span}\{f\} \setminus \{0\}))\\ &= \operatorname{posi}(\mathcal{L}_{\geq 0} \cup \mathcal{A}_0 \cup (\operatorname{span}\{f\} \setminus \{0\})),\\ &= \mathcal{A}_0 + \operatorname{span}\{f\}\end{aligned}$$

will also be coherent. Using Proposition 3$_{32}$, we know that

$$(\mathcal{I}_1, \mathcal{D}_1) \coloneqq (\mathcal{A}_1 \cap -\mathcal{A}_1, \mathcal{A}_1 \setminus -\mathcal{A}_1)$$

will also be coherent. Moreover, we know from Proposition 9$_{36}$ that the newly created couple $(\mathcal{I}_1, \mathcal{D}_1)$ will also respect $(\mathcal{I}_0, \mathcal{D}_0, \mathcal{I}_0 \cup \mathcal{D}_0)$ and therefore also the assessment $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$.

It only remains to prove that $\mathcal{A}_0 \cup (\operatorname{span}\{f\} \setminus \{0\})$ indeed avoids sure loss. We know that $\mathcal{A}_0 \cap (\operatorname{span}\{f\} \setminus \{0\}) = \emptyset$ whence it follows from Axiom A3$_{32}$ and A1$_{32}$ that $(\mathcal{A}_0 + \mathcal{L}_{>0}) \cap (\operatorname{span}\{f\} \setminus \{0\}) = \emptyset$. Applying Equation (1) twice and using $-\operatorname{span}\{f\} \setminus \{0\} = \operatorname{span}\{f\} \setminus \{0\}$ and $-\mathcal{L}_{>0} = \mathcal{L}_{<0}$, yields

$$\mathcal{A}_0 + (\operatorname{span}\{f\} \setminus \{0\}) \cap \mathcal{L}_{<0} = \emptyset.$$

As $\mathcal{A}_0$ avoids sure loss we infer that

$$\underbrace{((\mathcal{A}_0 + (\operatorname{span}\{f\} \setminus \{0\})) \cup \mathcal{A}_0)}_{\mathcal{A}_1} \cap \mathcal{L}_{<0} = \emptyset,$$

and we conclude that $\mathcal{A}_1$, and therefore also $\mathcal{A}_0$, avoid sure loss. $\square$

Proposition 17: *Let the assessment* $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ *avoid sure loss. Then there is at least one resolved model respecting it.*

Proof: If the assessment $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ avoids sure loss, then it can be extended by Theorem 11_{37} to a coherent set of acceptable gambles $\mathcal{A}_0 := \operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ that respects the assessment. If we apply Lemma 16 k times, then we get a set of acceptable gambles

$$\mathcal{A}_k = \mathcal{A}_0 + \sum_{i=1}^{k} \operatorname{span}\{f_k\}.$$

If we choose the gambles f_k such that

$$f_k \notin \operatorname{span}\{f_i : i \in \{1, \dots, k-1\}\},$$

then the proof follows from the finite dimension of the state-space. □

Notice that this result works also on infinite $\mathcal{X}$ provided we adopt the Axiom of Choice.

The following theorem shows that the natural extension can be written as the lower envelope of the resolved models.

Theorem 18 (Lower envelope theorem): *An assessment* $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ *is extendable if and only if* $\mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})} \neq \emptyset$. *Moreover,*

$$\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}) = \bigcap \mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})}.$$

Proof: $\mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})} \neq \emptyset$, means that there is some dominating coherent set of acceptable gambles that respects the assessment. So this assessment is extendable by Theorems 7_{35} and 11_{37}.

If $(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ is extendable, then $\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}) \in \mathbb{A}$ by Theorem 11_{37} and it follows immediately from Proposition 17 that $\mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})} \neq \emptyset$.

Because $\bigcap \mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})} = \bigcap \mathbb{M}_{\sqsupseteq \operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})}$ it is clear that

$$\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}) \subseteq \bigcap \mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})}.$$

Let us assume *ex absurdo* that $\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}) \subset \bigcap \mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})}$ which means that we can find a gamble $f \in \bigcap \mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})} \setminus \operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$. We infer that:

1. For every resolved set $\mathcal{M} \in \bigcap \mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}})$, the gamble f is acceptable, $f \in \mathcal{M}$.
2. $f \notin \operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$ implies that $\operatorname{posi}\{f\} \cap \operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as}) = \emptyset$. From Axioms A2_{32} and A3_{32}, it follows then that

$$\emptyset = \operatorname{posi}\{f\} \cap [\mathcal{L}_{>0} + \operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})]$$

$$
\begin{aligned}
&\Updownarrow \text{ by Equivalence (1)}\\
&\emptyset = \mathcal{L}_{<0} \cap [-\operatorname{posi}\{f\} + \operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})]\\
&\Updownarrow \text{ by Equation (1.9)}\\
&\emptyset = \mathcal{L}_{<0} \cap \operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as} \cup \{-f\}, \mathcal{A}_{as}).
\end{aligned}
$$

This means that the newly created model $\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as} \cup \{-f\}, \mathcal{A}_{as})$ avoids sure loss and is by Theorem 11_{37} coherent. The model is moreover constructed in such a way that the gamble f is undesirable. We know from Proposition 17 that $\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as} \cup \{-f\}, \mathcal{A}_{as})$ can be extended to a resolved model $\mathcal{M}$ that respects $\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as} \cup \{-f\}, \mathcal{A}_{as})$. Consequently, $f \in -\mathcal{M} \setminus \mathcal{M}$ and $\mathcal{M}$ respects $\operatorname{ext}(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})$

In the second point we have constructed a resolved model $\mathcal{M}$ that respects the assessment and should therefore be in $\bigcap \mathbb{M}_{\sqsupseteq(\mathscr{I}_{as}, \mathcal{D}_{as}, \mathcal{A}_{as})}$. However, the resolved set $\mathcal{M}$ does not contain the gamble f which is in contradiction with the first point. □

1.3 Multivariate acceptability

Most interesting problems involve more than one random variable and in order to be a worthy contender, our theory needs to be able to deal with this. In order to show that it does, we adopt the elegant notation used by De Cooman and Miranda [17].

In the most general case, we will consider a finite number of logically independent random variables X_N with $N \subset \mathbb{N}_{\geq 0}$, taking values in the respective sets $\mathcal{X}_n$, $n \in N$. With logically independent we mean that we can not a priori exclude values from the possibility space $\times_{n \in N} \mathcal{X}_n$. For every subset $R \subseteq N$, X_R denotes the tuple of variables taking values in the Cartesian product space $\mathcal{X}_R := \times_{r \in R} \mathcal{X}_r$ and elements of X_R will be denoted with lowercase letters x_R.

If $R = \emptyset$, then $\mathcal{X}_\emptyset$ contains by definition only one element $x_\emptyset := \emptyset$ whence $I_{X_\emptyset} = I_{\{x_\emptyset\}} = 1$. The set of all gambles on $\mathcal{X}_\emptyset$ is given by $\mathcal{L}(X_\emptyset) : \emptyset \to \mathbb{R}$ and can be identified with $\mathbb{R}$. There is only one coherent set of acceptable gambles on $X_\emptyset$: the set $\mathbb{R}_{\geq 0}$ of non-negative real numbers.

Following the interpretation given to a set of acceptable gambles in Section $1.2.1_{29}$, the specification of beliefs about the variables X_N involves the classification of the gambles $f \in \mathcal{L}(X_N)$ on the product space $\mathcal{X}_N$. As before $f(x_N)$ will be the reward for the gamble $f \in \mathcal{L}(X_N)$ if $X_N = x_N$ or in other words, if $X_n = x_n$ for all $n \in N$. Nothing essential changes, but instead of working with the space $\mathcal{X}$, the product space $\mathcal{X}_N$ is used.

1.3.1 Marginalising

What does a coherent set of acceptable gambles[4] $\mathcal{A}_{X_N}$ for the variables X_N tell us about beliefs about X_R alone, where $R \subset N$? Which will be the gambles on $\mathcal{L}(X)$ that are acceptable, desirable or indifferent? We reason that a gamble that does not depend on the value of $X_{N \setminus R}$ might as well be considered as a gamble on $\mathcal{L}(X_R)$ because the effect of $X_{N \setminus R}$ is nil. By introducing the projection operator proj_R as

$$\mathrm{proj}_R \colon \mathcal{X}_N \to \mathcal{X}_R \colon \mathrm{proj}_R(x_N) = x_R, \tag{1.14}$$

we can write such a gamble as $f \circ \mathrm{proj}_R$ where f is some gamble on $\mathcal{X}_R$.[5] The assumption we now make is that a gamble f on $\mathcal{L}(X_R)$ is acceptable if the gamble $f \circ \mathrm{proj}_R$ on $\mathcal{L}(X_N)$ is considered acceptable and we will also use the simpler notation $f \in \mathcal{A}$ instead of $f \circ \mathrm{proj}_R \in \mathcal{A}$. So we implicitly identify the gamble $f \in \mathcal{L}(X_R)$ and the gamble $f \circ \mathrm{proj}_R \in \mathcal{L}(X_N)$, and we also identify $\mathcal{L}(X_R)$ and $\mathrm{proj}_R^T(\mathcal{L}(X_R))$, where we let

$$\mathrm{proj}_R^T \colon \mathcal{L}(X_R) \to \mathcal{L}(X_N) \colon \mathrm{proj}_R^T(f) = f \circ \mathrm{proj}_R, \tag{1.15}$$

Definition 19 (Marginal): *The X_R-**marginal** $\mathrm{marg}_R(\mathcal{A}) \subseteq \mathcal{L}(X_R)$ of the set of acceptable gambles $\mathcal{A}$ on $\mathcal{L}(X_N)$ is given by*

$$\begin{aligned} \mathrm{marg}_R(\mathcal{A}) &\coloneqq \mathcal{L}(X_R) \cap \mathcal{A}, \\ &= \{f \in \mathcal{L}(X_R) \colon f \circ \mathrm{proj}_R \in \mathcal{A}\}, \\ &= \left(\mathrm{proj}_R^T\right)^{-1}(\mathcal{A}). \end{aligned}$$

Observe that $\mathrm{marg}_{\emptyset}(\mathcal{A}) = \mathbb{R}_{\geq 0}$ if $\mathcal{A}$ avoids sure loss and $\mathbb{R}$ otherwise. The following proposition (see also for example Equation (4) in [17]) is a simple consequence of the definition.

---

[4] If it is clear from the context, then we write $\mathcal{A}$ instead of $\mathcal{A}_{X_N}$.

[5] In Walley's terminology [62, par. 4.3.4, par. 6.2.5] for two variables X_1 and X_2, we say that the gambles $f \circ \mathrm{proj}_{\{1\}}$ are $\mathcal{B}$-measurable where the partition $\mathcal{B}$ is defined by the projection map $\mathrm{proj}_{\{1\}}$ as $\mathcal{B} \coloneqq \left\{x \in \mathcal{X}_1 \times \mathcal{X}_2 \colon \mathrm{proj}_{\{1\}}(x) \in \mathcal{X}_1\right\} = \{\{x_1\} \times \mathcal{X}_2 \colon x_1 \in \mathcal{X}_1\}$.

Proposition 20: *For any set of acceptable gambles $\mathcal{A}$ on $\mathcal{L}(X_N)$ and sets $V \subseteq R \subseteq N$:*

$$\operatorname{marg}_V(\operatorname{marg}_R(\mathcal{A})) = \operatorname{marg}_V(\mathcal{A})$$

The coherence of a marginal is a direct consequence of the coherence of the original model.

Proposition 21: *If a set of acceptable gambles $\mathcal{A}_{X_N}$ on $\mathcal{X}_N$ is coherent, then its X_R-marginal $\mathcal{A}_{X_R} = \operatorname{marg}_R\left(\mathcal{A}_{X_N}\right)$ will be coherent as well. Moreover the corresponding sets of indifferent gambles $\mathcal{I}_{X_R} := \mathcal{A}_{X_R} \cap -\mathcal{A}_{X_R}$ and desirable gambles $\mathcal{D}_{X_R} := \mathcal{A}_{X_R} \setminus -\mathcal{A}_{X_R}$ are given by*

$$\mathcal{I}_{X_R} = \mathcal{L}(X_R) \cap \mathcal{I}_{X_N} = \left\{f \in \mathcal{L}(X_R) : f \circ \operatorname{proj}_R \in \mathcal{I}_{X_N}\right\} = \operatorname{marg}_R\left(\mathcal{I}_{X_N}\right),$$
$$\mathcal{D}_{X_R} = \mathcal{L}(X_R) \cap \mathcal{D}_{X_N} = \left\{f \in \mathcal{L}(X_R) : f \circ \operatorname{proj}_R \in \mathcal{D}_{X_N}\right\} = \operatorname{marg}_R\left(\mathcal{D}_{X_N}\right).$$

Proof: Let us first remark that the following properties hold for the projection operator proj_R

$$f \,\Box\, g \Leftrightarrow f \circ \operatorname{proj}_R \,\Box\, g \circ \operatorname{proj}_R, \tag{1.16}$$
$$\alpha(f \circ \operatorname{proj}_R) + (g \circ \operatorname{proj}_R) = (\alpha f + g) \circ \operatorname{proj}_R, \tag{1.17}$$

with $\Box \in \{\leq, <, \prec, \succ, >, \geq\}$ and $\alpha \in \mathbb{R}$ and f and g any gambles in $\mathcal{L}(X_R)$.

Using these properties it is now straightforward to show that if $\mathcal{A}_{X_N}$ satisfies axioms $\mathrm{A1}_{32}$, $\mathrm{A2}_{32}$ and $\mathrm{A3}_{32}$, then so does $\mathcal{A}_{X_R}$. □

In the inverse problem to marginalization, we wonder what the implications of marginal beliefs are on a larger space; this is a special case of the natural extension problem addressed at the end of the previous section. Basically, if a marginal model $\mathcal{A}_{X_R}$ is given and it has to be extended to a larger space $\mathcal{X}_N$, then the natural extension is assumed to be the correct tool. Moreover, we know that avoiding sure loss of the marginal will guarantee a coherent joint model $\mathcal{A}_{X_N}$.

The natural extension ext_{X_N} can be explicitly written as

$$\operatorname{ext}_{X_N}(\mathcal{A}_{X_R}) = \operatorname{posi}(\mathcal{A}_{X_R} \cup \mathcal{L}_{\geq 0}(X_N)),$$

where we have identified the sets $\mathcal{A}_{X_R}$ and $\left(\operatorname{proj}_R^T\right)^{-1}(\mathcal{A}_{X_R})$. It is clear from their definition that $\operatorname{marg}_R\left(\operatorname{ext}_{X_N}(\mathcal{A}_{X_R})\right) = \mathcal{A}_{X_R}$ whenever $\mathcal{A}_{X_R}$ is coherent. The converse however does not hold. The strongest that can be said is that $\operatorname{ext}_{X_N}\left(\operatorname{marg}_R\left(\mathcal{A}_{X_N}\right)\right) \subseteq \mathcal{A}_{X_N}$. This is because the information about the relation between X_R and $X_{N\setminus R}$ that was lost by marginalising to $\mathcal{X}_R$ cannot be recovered.

To end, we give the following property for the marginal of a resolved model.

Proposition 22: *The marginal* $\operatorname{marg}_R(\mathcal{M})$ *of a resolved model* $\mathcal{M} \in \mathbb{M}_{X_N}$ *is again resolved:*

Proof: Assume that $f \notin \operatorname{marg}_R(\mathcal{M})$. This means that $\operatorname{proj}_R^T(f) \notin \mathcal{M}$ and then we know by Proposition 1539 that $-\operatorname{proj}_R^T(f) \in \mathcal{M}$. But then it follows by Definition 19 43 that $-\operatorname{proj}_R^T(f) = \operatorname{proj}_R^T(-f)$, and therefore $\operatorname{marg}_R(\mathcal{M})$ is resolved by Proposition 15 39. □

1.3.2 Conditioning

"What gambles will remain or become acceptable after observing an event?", is the question of **updating**. It is a special type of conditioning where it is known that the outcome of the random variable is confined to a subset of the possibility space. We will focus on the special events where the outcomes of a collection of random variables are known, i.e. $X_R = x_R$. If we are given a set of acceptable gambles $\mathcal{A}_{X_N}$ and observe this event $X_R = x_R$, then the only uncertainty that still remains, and needs to be modelled, concerns the variables $X_{N\setminus R}$, so the updated set $\mathcal{A}\rfloor x_R$ should be a subset of $\mathcal{L}(X_{N\setminus R})$. We postulate that the restriction of an acceptable called-off gamble is acceptable: For any gamble f in $\mathcal{L}(X_{N\setminus R})$ we have that

$$f \in \mathcal{A}\rfloor x_R \Leftrightarrow I_{x_R} f \in \mathcal{A}. \tag{1.18}$$

A gamble contingent on an event is the gamble that is called off unless the event occurs. This is what Walley calls the **updating principle** [62, Sec. 6.1.6] and what we refer to as the **contingent updating interpretation**. We define the **cylindrical extension** as

$$\operatorname{cylext}_{x_R}: \mathcal{X}_{N\setminus R} \to \mathcal{X}_N : \operatorname{cylext}_{x_R}(x_{N\setminus R}) = (x_R, x_{N\setminus R}).$$

Its lifted and its inverse variant are

$$\operatorname{cylext}_{x_R}^T : \mathcal{L}(\mathcal{X}_N) \to \mathcal{L}(\mathcal{X}_{N\setminus R}) : \operatorname{cylext}_{x_R}^T(f) = f \circ \operatorname{cylext}_{x_R} = f(x_R, \cdot),$$
$$\operatorname{cylext}_{x_R}^{-1} : \mathcal{L}(\mathcal{X}_{N\setminus R}) \to 2^{\mathcal{L}(\mathcal{X}_N)} : \operatorname{cylext}_{x_R}^{-1}(f) = \{h \in \mathcal{L}(\mathcal{X}_N) : h(x_R, \cdot) = f\}.$$

Following Walley and Moral, we can introduce the conditional model $\mathcal{A}|x_R$ which corresponds to the set of gambles on X_N that are acceptable contingent on $X_R = x_r$. We also allow conditioning on X_R with $R = \varnothing$. Since $I_{\mathcal{X}_\varnothing} = 1$, this is the same as not conditioning at all.

Geometrically, updating is tantamount to taking the intersection of the set of acceptable gambles with the subspace spanned by the indicators of the singletons of the conditioning event: $\mathcal{A}\rfloor x_R = \mathrm{cylext}^T_{x_R}(I_{x_R}\mathcal{A} \cap \mathcal{A})$. Conditioning involves an extra cylindrical extension $\mathcal{A}|x_R = \mathrm{cylext}^{-1}_{x_R}(\mathcal{A}\rfloor x_R)$. So for all $f \in \mathcal{L}(X_N)$

$$\begin{aligned} f(x_R,\cdot) \in \mathcal{A}\rfloor x_R &\Leftrightarrow I_{x_R} f(x_R,\cdot) \in \mathcal{A} \\ &\Leftrightarrow I_{x_R} f \in \mathcal{A} \Leftrightarrow f \in \mathcal{A}|x_R. \end{aligned} \tag{1.19}$$

We will always assume that we start with a coherent set of acceptable gambles and clearly we would like this coherence to be transferred to the updated set. There are however, special situations where the presented updating rule (1.18) produces incoherent sets.

Lemma 23 (Coherence of the updated set): *Consider a non-empty event $X_R = x_R$ and a coherent set $\mathcal{A}$ of acceptable gambles on $\mathcal{L}(X_N)$, then $\mathcal{A}\rfloor x_R$ satisfies* A2*32* *and* A3*32*. *Moreover, the following statements are equivalent:*

(i) $\mathcal{A}\rfloor x_R \cap \mathcal{L}_{<0}(X_{N\setminus R}) = \varnothing$, *i.e.* $\mathcal{A}\rfloor x_R$ *satisfies* A1*32*;

(ii) $\mathcal{A}\rfloor x_R \neq \mathcal{L}(X_{N\setminus R})$;

(iii) $\mathcal{A}\rfloor x_R$ *is a coherent set of acceptable gambles on* $\mathcal{L}(X_{N\setminus R})$;

(iv) $-I_{x_R} \notin \mathcal{A}$;

(v) $I_{x_R} \notin \mathcal{I}$, *where* $\mathcal{I} = \mathcal{A} \cap -\mathcal{A}$.

Proof: It is obvious that $\mathcal{A}\rfloor x_R$ satisfies A2$_{32}$ and A3$_{32}$, since $\mathcal{A}$ does. As a result, $\mathcal{A}\rfloor x_R = \mathrm{posi}(\mathcal{L}_{\geq 0}(x_R) \cup \mathcal{A}\rfloor x_R)$, and we infer from Theorem 7$_{35}$ that (i)–(iii) are equivalent. Obviously, (iv) and (v) are equivalent because $I_{x_R} \in \mathcal{A}$ [use $I_{x_R} \geq 0$ and A2$_{32}$]. It therefore remains to show that $\mathcal{A}\rfloor x_R$ satisfies A1$_{32}$ if and only if $-I_{x_R} \notin \mathcal{A}$.

Assume that $-I_{x_R} \in \mathcal{A}$. We show that $\mathcal{A}\rfloor x_R$ does not satisfy A1$_{32}$. Indeed, we infer from $-I_{x_R} \in \mathcal{A}$ and Equation (1.18) that the gamble -1 belongs to $\mathcal{A}\rfloor x_R$. Hence $\mathcal{A}\rfloor x_R$ does not satisfy A1$_{32}$.

Conversely, assume that $\mathcal{A}\rfloor x_R$ does not satisfy A1$_{32}$. This means that there is some $g \in \mathcal{A}\rfloor x_R$ such that $s := \max g < 0$. By Equation (1.18), $I_{x_R} g \in \mathcal{A}$, and therefore

$I_{x_R}s = I_{x_R}g + I_{x_R}[s-g] \in \mathcal{A}$, using A2$_{32}$ and A3$_{32}$. Use A3$_{32}$ one more time to find that $-I_{x_R} \in \mathcal{A}$. □

The situation that leads to incoherence corresponds to updating on an event considered to be impossible. We call an observation $X_R = x_r$ such that $I_{x_R} \in \mathcal{I}$ **practically impossible**. If no extra information is given, then we assume that the rational thing to do when updating on a practically impossible event, is to assume the vacuous model (see Example 12$_{37}$) for $\mathcal{A}\rfloor x_R$.

Definition 24: *Given a model $\mathcal{A} \subseteq \mathcal{L}(\mathcal{X}_N)$ for X_N and an event $X_R = x_R$, then the conditional model $\mathcal{A}|x_R \subseteq \mathcal{L}(X_N)$ and the updated model $\mathcal{A}\rfloor x_R := \mathcal{A}|x_R \cap \mathcal{L}(X_{N\setminus R})$ are given by*

$$\mathcal{A}|x_R := \begin{cases} \operatorname{cylext}_{x_R}^{-1}(I_{x_R}\mathcal{A} \cap \mathcal{A}) & \text{when } I_{x_R} \notin \mathcal{I}, \\ \mathcal{L}_{\geq 0}(\mathcal{X}_N) & \text{otherwise,} \end{cases}$$

$$\mathcal{A}\rfloor x_R = \begin{cases} \operatorname{cylext}_{x_R}^{T}(I_{x_R}\mathcal{A} \cap \mathcal{A}) & \text{when } I_{x_R} \notin \mathcal{I}, \\ \mathcal{L}_{\geq 0}(X_{N\setminus R}) & \text{otherwise.} \end{cases}$$

For any gamble $f \in \mathcal{L}(X_{N\setminus R})$, this can be formulated as

$$\begin{aligned} f \in \mathcal{A}\rfloor x_R &\Leftrightarrow [(I_{x_R}f \in \mathcal{A}) \wedge (-I_{x_R} \notin \mathcal{A})] \vee f \geq 0 \\ &\Leftrightarrow [(I_{x_R}f \in \mathcal{A}) \wedge (I_{x_R} \notin \mathcal{I})] \vee f \geq 0. \end{aligned} \tag{1.20}$$

We could have saved ourselves some trouble by making the rationality axioms stricter. In particular, if we had chosen $\mathcal{L}_{\geq 0} \subseteq \mathcal{D}$ as a rationality axiom instead of accepting sure gain (Axiom ID2$_{30}$), then the updating rule would have been $\mathcal{A}\rfloor x_R := \operatorname{cylext}_{x_R}^{T}(I_{x_R}\mathcal{A} \cap \mathcal{A})$ without extra conditions. The resulting set of rationality axioms would be stronger and less expressive than the ones presented. The newly created model would also force an elicitor to avoid assessments that correspond to a judgement of practical impossibility, whereas the model given here deals with these practical impossibilities by recognising them and resetting the beliefs when observing a practically impossible event [62, §2.1.2].

The question naturally arises whether updating on $X_R = x_R$ with $R \subseteq N$ and then updating on $X_V = x_V$ with $V \subseteq N$ is the same as updating on $X_{R\cup V} = x_{R\cup V}$. The following proposition shows that the order of updating is of no importance.

Proposition 25 (commutativity of updating): *Consider a collection of random variables X_N and a coherent set of acceptable gambles $\mathcal{A} \subseteq \mathcal{L}(X_N)$. Given two disjoint, nonempty sets $T \subseteq N$ and $V \subseteq N$, it holds that*

$$\mathcal{A}\rfloor x_{R\cup V} = (\mathcal{A}\rfloor x_R)\rfloor x_V = (\mathcal{A}\rfloor x_V)\rfloor x_R.$$

Proof: Using Equation (1.20) we see that for any $f \in \mathcal{L}(X_{N\setminus(R\cup V)})$

$$f \in (\mathcal{A}\rfloor x_R)\rfloor x_V \Leftrightarrow [\underbrace{(I_{x_V} f \in \mathcal{A}\rfloor x_R)}_{(A)} \wedge \underbrace{(-I_{x_V} \notin \mathcal{A}\rfloor x_R)}_{(B)}] \vee (f \geq 0). \tag{1.21}$$

Notice that $I_{x_V} \in \mathcal{L}(X_{N\setminus R})$. We develop statements (A) and (B) separately using Equation (1.20)

$$\begin{aligned} I_{x_V} f \in \mathcal{A}\rfloor x_R &\Leftrightarrow [I_{x_{R\cup V}} f \in \mathcal{A}) \wedge (-I_{x_R} \notin \mathcal{A})] \vee (I_{x_V} f \geq 0) \\ &\Leftrightarrow [I_{x_{R\cup V}} f \in \mathcal{A}) \wedge (-I_{x_R} \notin \mathcal{A})] \vee (f \geq 0), \\ -I_{x_V} \notin \mathcal{A}\rfloor x_R &\Leftrightarrow (-I_{x_{R\cup V}} \notin \mathcal{A}) \vee (-I_{x_R} \in \mathcal{A}). \end{aligned}$$

By substituting these expressions back into Equation (1.21), and by putting the logical expression in its conjunctive normal form, we get

$$\begin{aligned} f \in (\mathcal{A}\rfloor x_R)\rfloor x_V \Leftrightarrow\ & [(I_{x_{R\cup V}} f \in \mathcal{A}) \wedge (-I_{x_R} \notin \mathcal{A}) \wedge (-I_{x_{R\cup V}} \notin \mathcal{A})] \\ & \vee [(I_{x_{R\cup V}} f \in \mathcal{A}) \wedge (-I_{x_R} \notin \mathcal{A}) \wedge (-I_{x_R} \in \mathcal{A})] \\ & \vee [(f \geq 0) \wedge (-I_{x_{R\cup V}} \notin \mathcal{A})] \vee [(f \geq 0) \wedge (-I_{x_R} \in \mathcal{A})] \\ & \vee (f \geq 0) \\ \Leftrightarrow\ & [(I_{x_{R\cup V}} f \in \mathcal{A}) \wedge (-I_{x_R} \notin \mathcal{A}) \wedge (-I_{x_{R\cup V}} \notin \mathcal{A})] \vee (f \geq 0). \end{aligned}$$

The last equivalence can be simplified. Because $I_{x_R} \geq I_{x_{R\cup V}}$, it follows from Lemma 2_{31} and A2_{32} that if $-I_{x_R} \in \mathcal{A}$ then $-I_{x_{R\cup V}} \in \mathcal{A}$. We thus get that

$$\begin{aligned} f \in (\mathcal{A}\rfloor x_R)\rfloor x_V &\Leftrightarrow [(I_{x_{R\cup V}} f \in \mathcal{A}) \wedge (-I_{x_{R\cup V}} \notin \mathcal{A})] \vee (f \geq 0) \\ &\Leftrightarrow f \in \mathcal{A}\rfloor x_{R\cup V}. \end{aligned}$$ □

It can be shown rather easily that the updating rule keeps indifferent gambles indifferent and desirable gambles desirable in the conditioning subspace, under the assumption that the updating event is not practically impossible.

Proposition 26: *Given a model $(\mathcal{I}, \mathcal{D})$ for X_N and an event $X_R = x_R$, then the updated sets $\mathcal{I}\rfloor x_R := (\mathcal{A}\rfloor x_R) \cap (-\mathcal{A}\rfloor x_R)$ and $\mathcal{D}\rfloor x_R := (\mathcal{A}\rfloor x_R) \setminus (-\mathcal{A}\rfloor x_R)$ are given by*

$$\mathcal{I}\rfloor x_R = \{f \in \mathcal{L}(X_{N\setminus R}) \colon I_{x_R} f \in \mathcal{I} \text{ and } I_{x_R} \notin \mathcal{I}\} \cup \{0\},$$

$$\mathcal{D}\rfloor x_R = \{f \in \mathcal{L}(X_{N\setminus R})\colon I_{x_R} f \in \mathcal{D} \text{ and } I_{x_R} \notin \mathcal{I}\} \cup \mathcal{L}_{>0}(x_{N\setminus R}).$$

The updated model is defined on the space $\mathcal{X}_{N\setminus R}$, but if a gamble on the complete space $\mathcal{X}_N$ is provided and if we know for some reason that the outcome will assume $X_R = x_R$, then the conditional model $\mathcal{A}|x_R$ can be used. This model does not care about the gamble's values in $\{x_R\}^c \times \mathcal{X}_{N\setminus R}$.

If, on the other hand, the conditional model is considered to be a local assessment for a greater space, then it is a bit presumptuous to assume that the updating event will certainly take place. The contingent updating interpretation dictates that the only assumption we want to make in this case is that the extended and the updated models are equal in the hyperplane defined by the updating event $X_R = x_r$. In other words, given a conditional model $\mathcal{A}\rfloor x_R$ with $R \subset N$, then the most conservative coherent extension to $\mathcal{L}(X_N)$ is given by $\mathrm{ext}(\mathcal{A}\rfloor x_R)$. When the updating event is not practically impossible, this extension is equal to $\mathrm{ext}(I_{x_R}\mathcal{A} \cap \mathcal{A})$ and is also called the **weak extension** by Moral [42].

1.3.3 Combining partial models

We see from Definitions 19_{43} and 24_{47} that—apart from a degenerate situation where the updating event lies in the boundary of the cone of desirable gambles—both marginalising and updating amount to taking intersections of the set of acceptable gambles with a hyperplane: $\mathcal{L}(\mathcal{X}_R)$ when marginalising to $\mathcal{X}_R$ and $I_{\{x_R\}\times\mathcal{X}_{N\setminus R}}\mathcal{L}(\mathcal{X}_N)$ when updating on $X_R = x_R$. Intuitively we know that the order in which intersections are taken is irrelevant, which means that the marginal of the conditional is exactly the same as the conditional of a marginal. This is made explicit in the following proposition whose counterpart for sets of desirable gambles was proved by Moral [42] and de Cooman, Miranda & Zaffalon [17, Proposition 9]. The proof of the proposition follows the idea behind the discussion in the latter paper.

Proposition 27: *For a coherent set of acceptable gambles $\mathcal{A} \in \mathcal{L}(X_N)$ and two disjoint sets $V \subseteq N$ and $R \subseteq N$ it holds that*

$$\mathrm{marg}_{R\cup V}(\mathcal{A})\rfloor x_R = \mathrm{marg}_V(\mathcal{A}\rfloor x_R).$$

Proof: Let us first remark that $I_{x_R} \in \mathcal{I} := \mathcal{A} \cap -\mathcal{A}$ is equivalent to $I_{x_R} \in \mathcal{A} \cap -\mathcal{A} \cap \mathcal{L}(X_{R\cup V}) = \operatorname{marg}_{R\cup V}(\mathcal{A}) \cap -\operatorname{marg}_{R\cup V}(\mathcal{A})$ because $I_{x_R} \in \mathcal{L}(X_{R\cup V})$ and $-\mathcal{L}(X_{R\cup V}) = \mathcal{L}(X_{R\cup V})$.

Consider any $f \in \mathcal{L}(X_V)$ and observe the following chain of equivalences:

$$
\begin{aligned}
f \in \operatorname{marg}_{R\cup V}(\mathcal{A})\rfloor x_R &\Leftrightarrow [(I_{x_R} f \in \operatorname{marg}_{R\cup V}(\mathcal{A})) \wedge (I_{x_R} \notin \operatorname{marg}_{R\cup V}(\mathcal{I}))] \vee (f \geq 0) \\
&\Leftrightarrow [(I_{x_R} f \in \mathcal{A}) \wedge (I_{x_R} \notin \mathcal{I})] \vee (f \geq 0) \\
&\Leftrightarrow f \in \mathcal{A}\rfloor x_R
\end{aligned}
$$

and since $f \in \mathcal{L}(X_V)$, this is equivalent with $f \in \operatorname{marg}_V(\mathcal{A}\rfloor x_R)$. □

The question to address now, is how to combine local models, i.e. conditional and marginal models, into a joint model. A **local model** that carries the modeller's belief about X_V after updating on $X_R = x_R$ will be denoted by $\mathcal{A}_{V\rfloor x_R}$. When each of the local models is coherent, we say that the local models are separately coherent.

Definition 28: *Consider a collection of local models $\{\mathcal{A}_{V_i\rfloor x_{R_i}}\}_{i\in I}$ with I any index set, then we say that the local models are* ***separately coherent*** *if and only if each $\mathcal{A}_{V_i\rfloor x_{R_i}}$ in the collection is a coherent set of acceptable gambles in $\mathcal{L}(X_{V_i})$.*

Before moulding local models into a joint model $\mathcal{A}$, we have to consider the behavioural consequences for each local model $\mathcal{A}_{V\rfloor x_R}$ on the joint space $\mathcal{X}_N$. For example, if a (marginal) gamble $f \in \mathcal{L}(X_V)$ is considered acceptable, then each gamble in $\operatorname{proj}_N^T(f)$ is acceptable in $\mathcal{L}(X_N)$ as well by Definition 1943. Similarly, if a gamble in $\mathcal{L}(X_{N\setminus R})$ is considered acceptable after updating on $X_R = x_R$, then the gamble $I_{x_R} f$ is by Definition 2447 acceptable in $\mathcal{L}(X_N)$. As we know from Proposition 27 that the order in which we update and marginalise is irrelevant, we infer that whenever a gamble f belongs to a local model $\mathcal{A}_{V\rfloor x_R}$, then the gambles $I_{x_R} \operatorname{proj}_{N\setminus R}^T(f)$ are acceptable in $\mathcal{L}(X_N)$ as well, where we assume that $N \supseteq V \cup R$.

Once the local models are reinterpreted on the joint domain, the natural extension of the union of these reinterpreted local models can be computed to get the most conservative model $\mathcal{A}$ that represents the behavioural consequences of the local models on the joint space. Unfortunately, this natural extension however is not guaranteed to be a coherent set of acceptable gambles, even if the local models are separately coherent. This is because the combination of different local models can incur sure loss in which case no coherent extension exists. But, even if this natural extension $\mathcal{A}$ is coherent, it is not guaranteed that every local model $\mathcal{A}_{V\rfloor x_R}$ is actually equal to $\mathrm{marg}_V(\mathcal{A}\rfloor x_R)$. This can for instance occur when the event $X_R = x_R$ turns out to be practically impossible in which case this apparent incompatibility is not a problem, on the contrary, the model $\mathcal{A}_{V\rfloor x_R}$ should then be considered as extra information, as a refinement of the modeller's belief. If the discrepancy between the model derived from the joint and the local model is not due to the conditioning on a practical impossible event, then this means that the modeller specified too stringent a local model. If the natural extension is coherent, then this local model can be extended, but the modeller should at least be informed whether that is really what he wants.

Definition 29: *The joint $\mathcal{A}$ of a collection of separately coherent local models $\{\mathcal{A}_{V_i\rfloor x_{R_i}}\}_{i\in I}$ on a joint domain $\mathcal{X}_N$ with $\bigcup_{i\in I}(R_i\cup V_i)\subseteq N$, is given by*

$$\mathcal{A} := \mathrm{ext}\left(\bigcup_{i\in I} I_{x_{R_i}}\,\mathrm{proj}^T_{N\setminus R_i}(\mathcal{A}_{V_i\rfloor x_{R_i}})\right).$$

If the joint $\mathcal{A}$ is coherent and if it holds for every $i\in I$ that

$$\mathcal{A}_{V_i\rfloor x_{R_i}} = \mathrm{marg}_{V_i}(\mathcal{A}\rfloor x_{R_i}) \text{ whenever } -I_{x_{R_i}} \notin \mathcal{A},$$

then we say that the local assessments $\mathcal{A}_{V_i\rfloor x_{R_i}}$ are ***jointly coherent****.*

As explained before, in general nothing can be said a priori about the coherence of the joint model, let alone about joint coherence of the local models. There are, however, specific cases where statements about joint coherence can be made on beforehand. One such a situation is marginal extension, which is the acceptability counterpart of the Towering Equality, or the Law of Iterated Expectation, in classical probability theory $\mathrm{P}_{X,Y}(\cdot) = \mathrm{P}_X(\mathrm{P}_Y(\cdot|X))$.

Theorem 30 (Marginal extension): *Consider a coherent set of acceptable gambles $\mathcal{A}_X$ on $\mathcal{L}(X)$. Consider moreover an updated coherent set of acceptable gambles $\mathcal{A}_{Y\rfloor x}\subseteq\mathcal{L}(Y)$ for every $x\in\mathcal{X}$. Then local models $\mathcal{A}_X$ and $\mathcal{A}_{Y\rfloor x}$ are jointly coherent, and their joint $\mathcal{A}\subseteq\mathcal{L}(X,Y)$ called the* ***marginal extension****, is given by*

$$\mathcal{A} := \sum_{x\in\mathcal{X}} I_{\{x\}}\mathcal{A}_{Y\rfloor x} + \mathrm{proj}^T_X(\mathcal{A}_X)$$

$$= \sum_{x\in\mathcal{X}} I_{\{x\}}\mathcal{A}_{Y\rfloor x} + \mathcal{A}_X,$$

Proof: We start by showing that $\mathcal{A}$ is coherent.

A1: Assume that $\mathcal{A}$ does not avoid sure loss. Then for some $x \in \mathcal{X}$ and $\epsilon > 0$ there are $h \in \mathcal{A}_X$ and $f_x \in \mathcal{A}_{Y\rfloor x}$ such that $\sum_{x\in\mathcal{X}} I_{\{x\}} f_x + \mathrm{proj}_X^T(h) < -\epsilon$. Since $h \in \mathcal{A}_X$ and $\mathcal{A}_X$ avoids sure loss, we know that $\max h \geq 0$. So there is some $x^* \in \mathcal{X}$ such that $h(x^*) \geq 0$, where $f_{x^*}(y) + h(x^*) < -\epsilon$ and therefore

$$f_{x^*}(y) < -\epsilon - h(x^*) \text{ for all } y \in \mathcal{Y}$$

But this would mean that $f_{x^*} < 0$ which contradicts the assumption that $\mathcal{A}_{Y\rfloor x^*}$ avoids sure loss.

A2: As all local models accept partial gain, we see that $\sum_{x\in X} I_{\{x\}}\mathcal{L}_{\geq 0}(Y) + \mathcal{L}_{\geq 0}(X) = \mathcal{L}_{\geq 0}(X,Y) \subseteq \mathcal{A}$.

A3: Assume that f and h belong to $\mathcal{A}$. Then we know that there are gambles $f_x, h_x \in \mathcal{A}_{Y\rfloor x}$ and $g_f, g_h \in \mathcal{A}_X$ such that

$$f = \sum_{x\in\mathcal{X}} I_{\{x\}} f_x + \mathrm{proj}_X^T(g_f),$$
$$h = \sum_{x\in\mathcal{X}} I_{\{x\}} h_x + \mathrm{proj}_X^T(g_h).$$

It follows from the separate coherence of the local models that for every $\alpha > 0$ and $\beta > 0$, $\alpha f_x + \beta h_x \in \mathcal{A}_{Y\rfloor x}$ and $\alpha g_f + \beta g_h \in \mathcal{A}_X$, whence also $\alpha f + \beta h \in \mathcal{A}$.

We conclude that $\mathcal{A}$ is indeed coherent and from Lemma 29 we also know that it is the least committal one.

To prove joint coherence we show first that the X-marginal of $\mathcal{A}$ equals $\mathcal{A}_X$. As $\mathrm{marg}_X(\mathcal{A}) = \mathcal{L}(X) \cap \mathcal{A}$ we look for gambles in $\sum_{x\in X} I_{\{x\}}\mathcal{A}_{Y\rfloor x} + \mathcal{A}_X$ that do not depend on Y. Clearly every gamble in $\mathcal{A}_X$ is in $\mathrm{marg}_X(\mathcal{A})$. Assume that $f \in \sum_{x\in\mathcal{X}} I_{\{x\}}\mathcal{A}_{Y\rfloor x}$ then $f \in \mathcal{L}(X)$ implies that $f(x, y_1) = f(x, y_2)$ for any y_1 and y_2 in $\mathcal{Y}$. This means that a constant gamble is picked from every $\mathcal{A}_{Y\rfloor x}$ and as $\mathcal{A}_{Y\rfloor x}$ avoids sure loss, we conclude that $f \geq 0$ and $\mathcal{A}_X + \{f\} \subseteq \mathcal{A}_X$ whence $\mathrm{marg}_X(\mathcal{A}) = \mathcal{A}_X$.

To finish, we must show that:

$$\mathcal{A}\rfloor x = \mathcal{A}_{Y\rfloor x} \text{ for all } x \text{ such that } -I_{\{x\}} \notin \mathcal{A}_X.$$

1. It is easy to show that $\mathcal{A}_{Y\rfloor x} \subseteq \mathcal{A}\rfloor x$. Indeed, $g \in \mathcal{A}_{Y\rfloor x}$ requires that $I_{\{x\}} g \in \mathcal{A}$ and by Definition 24$_{47}$ this implies that $g \in \mathcal{A}\rfloor x$.
2. Conversely, let $g \in \mathcal{A}\rfloor x$, so we know that $I_{\{x\}} g \in \mathcal{A}$ because $-I_{\{x\}} \notin \mathcal{A}_X$ and then $-I_{\{x\}} \notin \mathcal{A}$. So there are $h \in \mathcal{A}_X$ and $f_z \in \mathcal{A}_{Y\rfloor z}$, $z \in \mathcal{X}$ such that

$$I_{\{x\}} g = h + \sum_{z\in\mathcal{X}} I_{\{z\}} f_z.$$

Hence $g(\cdot) = h(x) + f_x(\cdot)$ and $0 = h(z) + f_z(\cdot)$ for all $z \in \mathcal{X} \setminus \{x\}$. For any $z \neq x$ we see that f_z is constant and therefore $f_z \geq 0$ whence $h(z) \leq 0$. This implies that $h(x) \geq 0$ because $h \in \mathcal{A}_X$ and $\mathcal{A}_X$ avoids sure loss. Since $g \geq f_x$ and $f_x \in \mathcal{A}_{Y\rfloor x}$ it follows that $g \in \mathcal{A}_{Y\rfloor x}$. □

1.4 Derived models

The $(\mathcal{I},\mathcal{D})$-model described so far is probably amongst the most expressive models capable of describing uncertainty. The downside of this expressive wealth is that it comes with serious computational problems that boil down to the problem that no current software library can do exact calculations with polyhedra that are partly open and partly closed.[6] To overcome these problems, the $(\mathcal{I},\mathcal{D})$-model can be relaxed a bit. One natural way to relax the conditions is by demanding all indifferent gambles to be desirable, or alternatively to be not desirable. This is what the the models in the upcoming sections partly do. Additional assumptions that reduce the expressiveness but improve the computability may be used. Of course one can think of different models that approach the $(\mathcal{I},\mathcal{D})$-model, at least one such model is the weak-desirability model given in [20]. We will restrict ourselves to natural and regular extension: the models most often referred to in the literature.

1.4.1 Almost-desirability

Almost-desirability was introduced by Walley [62, §3.7] and is closely related to regular extension [62, Appendix J]. We denote the set of almost desirable gambles by $\mathcal{R}$ and a gamble is said to be **almost desirable** if adding any positive amount of utility to it results in a desirable gamble:

$$f \in \mathcal{R} \Leftrightarrow (\forall \epsilon > 0)(f + \epsilon \in \mathcal{D}). \tag{1.22}$$

Lemma 31: *The set of **almost desirable** gambles $\mathcal{R}$ corresponding to $\mathcal{A}$ (or to $(\mathcal{I},\mathcal{D})$) is given by*

$$\mathcal{R} = \bigcap_{\epsilon>0}(\mathcal{D} - \epsilon) = \bigcap_{\epsilon>0}(\mathcal{A} - \epsilon).$$

[6] The Parma Polyhedra Library [2] can do calculations for what they call Nearly Closed Polyhedra which are polyhedra that have faces that are either open or closed. Cones of acceptable gambles that have faces that are partly open, partly closed cannot be modelled with this library.

Proof: From Equation (1.22) we know that $f \in \mathcal{R}$ if and only if $f \in \mathcal{D} - \epsilon$ for any $\epsilon > 0$ and therefore $\mathcal{R} = \bigcap_{\epsilon>0}(\mathcal{D} - \epsilon)$.

To prove the second equality, it is sufficient to show that $\bigcap_{\epsilon>0}(\mathcal{A} - \epsilon) \subseteq \bigcap_{\epsilon>0}(\mathcal{D} - \epsilon)$ since $\mathcal{D} \subseteq \mathcal{A}$ immediately leads to the inverse inclusion. Assume for a moment that this equality does not hold. Then there must be some gamble f such that

$$f \in \bigcap_{\epsilon>0}(\mathcal{A} - \epsilon) \text{ or equivalently } (\forall \delta > 0) f + \delta \in \mathcal{A},$$

and

$$f \notin \bigcap_{\epsilon>0}(\mathcal{D} - \epsilon) \text{ or equivalently } (\exists \epsilon > 0) f + \epsilon \notin \mathcal{D}.$$

Let $\delta < \epsilon$, then $\epsilon - \delta > 0$ and by ID3$_{30}$ also $\epsilon - \delta \in \mathcal{D}$. Then we infer from ID5$_{30}$ and ID7$_{30}$ that $f + \epsilon = \underbrace{f + \delta}_{\in \mathcal{A}} + \underbrace{\epsilon - \delta}_{\in \mathcal{D}} \in \mathcal{D}$, a contradiction. □

A set of almost desirable gambles has a number of properties that make it coherent as a set of acceptable gambles.

Proposition 32: *Let $\mathcal{A}$ be a coherent set of acceptable gambles on a space $\mathcal{L}$ and $\mathcal{R}$ the corresponding set of almost desirable gambles. Then*

AD1. $\mathcal{R} \cap \mathcal{L}_{<0} = \emptyset$ [avoiding sure loss]

AD2. $\mathcal{L}_{\geq 0} \subseteq \mathcal{R}$ [accepting partial gain]

AD3. $\text{posi}(\mathcal{R}) = \mathcal{R}$ [deduction]

AD4. $\bigcap_{\epsilon>0}(\mathcal{R} - \epsilon) = \mathcal{R}$ [closure[7]]

A set $\mathcal{R}$ that satisfies these conditions is said to be a coherent set of almost desirable gambles.

Proof: AD1: If $f < 0$ then there is some $\epsilon > 0$ such that $f + \epsilon < 0$. So by A1$_{32}$, $f + \epsilon \notin \mathcal{D}$, whence $f \notin \mathcal{R}$.

AD2: If $f \geq 0$ then by ID3$_{30}$ it holds for all $\epsilon > 0$ that $f + \epsilon \in \mathcal{D}$, whence $f \in \mathcal{R}$.

AD3: if $f_1 \in \mathcal{R}$ and $f_2 \in \mathcal{R}$ then for all $\epsilon > 0$ and $\alpha > 0$ there are $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $\epsilon = \alpha\epsilon_1 + \epsilon_2$ and $f_1 + \epsilon_1 \in \mathcal{D}$ and $f_2 + \epsilon_2 \in \mathcal{D}$. This implies by axiom ID5$_{30}$ that $\alpha f_1 + f_2 + \epsilon \in \mathcal{D}$ and therefore $\alpha f_1 + f_2 \in \mathcal{R}$.

AD4: $\bigcap_{\epsilon>0}(\mathcal{R} - \epsilon) = \bigcap_{\epsilon>0}\bigcap_{\delta>0}[\mathcal{D} - (\epsilon + \delta)] = \bigcap_{\epsilon>0}(\mathcal{D} - \epsilon) = \mathcal{R}$. □

[7] This axiom, taken together with AD1-AD3 makes sure that the cone $\mathcal{R}$ is closed in the usual Euclidean topology on the finite dimensional linear space $\mathcal{L}$.

If we compare the definition of coherence for sets of almost desirable gambles with that of coherent sets of acceptable gambles, we see that the only difference is the addition of the extra closure axiom AD4. From a mathematical perspective, a coherent set of almost desirable gambles is just a special type of coherent set of acceptable gambles. This means that we can take over most of the machinery developed for acceptability models; we only need to add an extra check for the closure axiom AD4.

▷ **Example 33:** The vacuous model is given by $\mathcal{R} = \mathcal{L}_{\geq 0}$.

If provided with a **finite** partial assessment $\mathcal{R}_{as}$, then nothing changes with respect to the acceptability case as the application of the posi operator will automatically produce a closed set. The only condition that needs to be checked in order for a partial assessment to be extendable to a set of almost desirable gambles is again the avoiding sure loss axiom AD1. In the most general case, with an infinite assessment, the extension has to take care of the closed character of the set of almost desirable gambles. The resulting extension is what we call the **regular extension** and is, for an assessment $\mathcal{R}_{as}$, given by

$$\operatorname{regext}(\mathcal{R}_{as}) \coloneqq \bigcap_{\epsilon>0} (\operatorname{ext}(\mathcal{R}_{as}) - \epsilon), \tag{1.23}$$

the topological closure of the natural extension as we know it.

The condition for a partial assessment $\mathcal{R}_{as}$ to be made coherent is again avoiding sure loss, which also means that there can only be a corresponding coherent set of acceptable gambles compatible with $\mathcal{R}$ if $\mathcal{R}$ avoids sure loss. One of the advantages is that for **finite** assessments, the cone of almost desirable gambles is now a finite intersection of half spaces and therefore by definition a closed polytope. This means that computations can be done using existing software packages for dealing with closed polytopes.

Updating sets of almost desirable gambles. The reasoning used for updating in the acceptability model remains the same when updating a set of almost desirable gambles. Gambles that were almost desirable are assumed to stay almost desirable after updating, unless this results in an incoherent set. As in the acceptability case, this will happen when the semispace spanned by the atoms of the updating event is a boundary of the cone of almost desirable gambles.

Proposition 34: *Given a coherent set of almost desirable gambles $\mathcal{R} \in \mathcal{L}(X_N)$ and an event $X_R = x_r$ with $R \subseteq N$, then the updated model $\mathcal{R}\rfloor x_R$ given by*

$$\mathcal{R}\rfloor x_R := \begin{cases} \operatorname{cylext}^T_{x_R}(I_{x_R}\mathcal{R} \cap \mathcal{R}) & \textit{when } -I_{x_R} \notin \mathcal{R}, \\ \mathcal{L}_{\geq 0}(X_{N\setminus R}) & \textit{otherwise,} \end{cases}$$

is a coherent set of almost desirable gambles. Alternatively,

$$f \in \mathcal{R}\rfloor x_R \Leftrightarrow [(I_{x_R} f \in \mathcal{R}) \wedge (-I_{x_R} \notin \mathcal{R})] \vee (f \geq 0).$$

Proof: As coherent sets of almost desirable gambles are coherent sets of acceptable gambles with the extra condition that Axiom AD4$_{54}$ should be fulfilled, we only need to check this last axiom.

To prove AD4$_{54}$ we infer from AD3$_{54}$, AD2$_{54}$ and Equation (1.22) that

$$f \in \mathcal{R}\rfloor x_R \Rightarrow (\forall \delta > 0) f + \delta \in \mathcal{R}\rfloor x_R \Leftrightarrow f \in \bigcap_{\delta>0} (\mathcal{R}\rfloor x_R - \delta), \text{ and}$$

$$\begin{aligned} f \in \bigcap_{\delta>0} (\mathcal{R}\rfloor x_R - \delta) &\Leftrightarrow (\forall \delta > 0) I_{x_R}(f + \delta) \in \mathcal{R} \\ &\Leftrightarrow (\forall \delta > 0)(\forall \epsilon > 0) I_{x_R}(f + \delta) + \epsilon \in \mathcal{R} \\ &\Rightarrow (\forall \gamma > 0) I_{x_R} f + \gamma \in \mathcal{R} \Leftrightarrow I_{x_R} f \in \mathcal{R} \Leftrightarrow f \in \mathcal{R}\rfloor x_R, \end{aligned}$$

where the unconditional version of Axiom AD4$_{54}$ was used to get the last implication. These implications prove that indeed $f \in \mathcal{R}\rfloor x_R \Leftrightarrow f \in \bigcap_{\delta>0}(\mathcal{R}\rfloor x_R - \delta)$. □

1.4.2 Strictly desirable gambles

As explained before, the set of almost desirable gambles is actually the topological closure of the set of desirable gambles. This means that in the worst case, some gambles might be called almost desirable that are actually undesirable. To prevent this situation, we could also opt for a simplified model that is a subset rather than a superset of the set of desirable gambles $\mathcal{D}$. The natural extension model we introduce here does just this by allowing only gambles that are **strictly desirable**, which means that

$$f \in \mathcal{N} \Leftrightarrow (\exists \epsilon > 0)(f - \epsilon \in \mathcal{A}) \Leftrightarrow (\exists \epsilon > 0)(f - \epsilon \in \mathcal{D}) \tag{1.24}$$

where $\mathcal{N}$ is the **set of strictly desirable gambles**. To show that the last equivalence holds it is sufficient to see that if there is an $\epsilon > 0$ such that $f - \epsilon \in \mathcal{A}$, then $f - \epsilon/2 \in \mathcal{A} + \epsilon/2 \subseteq \mathcal{D}$ by ID3$_{30}$, ID5$_{30}$ and ID7$_{30}$.

Proposition 35: *Let $\mathcal{A} \subset \mathcal{L}$ be a coherent set of acceptable gambles and $\mathcal{N}$ the corresponding set of strictly desirable gambles. Then*

SD1. $\mathcal{N} \cap \mathcal{L}_{\leq 0} = \emptyset$ [avoiding sure loss]

SD2. $\mathcal{L}_{>0} \subseteq \mathcal{N}$ [accepting sure gain]

SD3. $\operatorname{posi}(\mathcal{N}) = \mathcal{N}$ [deduction]

SD4. $\bigcup_{\epsilon>0}(\mathcal{N} + \epsilon) = \mathcal{N}$ [openness]

A set $\mathcal{N}$ that satisfies these conditions is called a coherent set of ***strictly desirable gambles****.*

Proof: From Equation (1.24) we infer that $f \notin \mathcal{N} \Leftrightarrow (\forall \epsilon > 0)(f - \epsilon \notin \mathcal{A}) \Leftrightarrow f \notin \bigcup_{\epsilon>0}(\mathcal{A} + \epsilon)$, whence

$$\mathcal{N} = \bigcup_{\epsilon>0}(\mathcal{A} + \epsilon). \tag{1.25}$$

SD1: For any $\epsilon > 0$ it holds by A1$_{32}$ that $\mathcal{A} \cap (\mathcal{L}_{\leq 0} - \epsilon) = \emptyset$ and this is by Equation (1) equivalent to $(\mathcal{A}+\epsilon) \cap \mathcal{L}_{\leq 0} = \emptyset$. Hence $\bigcup_{\epsilon>0}((\mathcal{A}+\epsilon) \cap \mathcal{L}_{\leq 0}) = \mathcal{L}_{\leq 0} \cap \bigcup_{\epsilon>0}(\mathcal{A} + \epsilon) = \emptyset$, and therefore $\mathcal{L}_{\leq 0} \cap \mathcal{N} = \emptyset$, by Equation (1.25).

SD2: If $f > 0$ then there is some $\epsilon > 0$ such that $f - \epsilon \geq 0$ whence $f - \epsilon \in \mathcal{D}$ by A2$_{32}$.

SD3: if $\alpha \geq 0$, $f_1 \in \mathcal{N}$ and $f_2 \in \mathcal{N}$ then there are $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $f_1 - \epsilon_1 \in \mathcal{D}$ and $f_2 - \epsilon_2 \in \mathcal{D}$ whence by ID5$_{30}$ $\alpha f_1 + f_2 + \alpha\epsilon_1 + \epsilon_2 \in \mathcal{D}$, which implies that $\alpha f_1 + f_2 \in \mathcal{N}$.

SD4: Using Equation (1.25) we infer that

$$\bigcup_{\epsilon>0}(\mathcal{N} + \epsilon) = \bigcup_{\epsilon>0}\left(\bigcup_{\delta>0}(\mathcal{A} + \delta) + \epsilon\right) = \bigcup_{\epsilon>0}(\mathcal{A} + \epsilon) = \mathcal{N}. \qquad \square$$

It is not possible to interpret a set of strictly desirable gambles $\mathcal{N}$ as a special type of sets of acceptable gambles. This is because SD4 and ID1$_{30}$ demand that $0 \notin \mathcal{N}$ which is in conflict with ID4$_{30}$.

Given a set of strictly desirable gambles $\mathcal{N}_{as}$, we can again calculate the natural extension $\operatorname{ext}(\mathcal{N}_{as})$, the smallest coherent set containing the assessment $\mathcal{N}_{as}$. This resulting set will only be coherent if the assessment avoids sure loss, i.e. if no gamble in $\operatorname{posi}(\mathcal{N}_{as})$ is in $\mathcal{L}_{\leq 0}$, and it is given by

$$\operatorname{natext}(\mathcal{N}_{as}) := \operatorname{posi}(\mathcal{N}_{as}) + \mathcal{L}_{>0}. \tag{1.26}$$

▷ **Example 36:** The vacuous model corresponds to $\mathcal{N} = \mathcal{L}_{>0}$.

The direct relation between coherent sets of strictly desirable gambles and coherent sets of almost desirable gambles is given in the following proposition.

Proposition 37: *Consider a coherent set of acceptable gambles $\mathcal{A}$ then the corresponding sets of almost desirable and strictly desirable gambles satisfy:*

$$\mathcal{R} = \bigcap_{\epsilon>0} (\mathcal{N} - \epsilon),$$
$$\mathcal{N} = \bigcup_{\delta>0} (\mathcal{R} + \delta).$$

Proof: From Lemma 31_{53}, Equation (1.24), $\mathrm{AD4}_{54}$ and SD4, we infer that

$$h \in \mathcal{R} \Leftrightarrow (\forall \epsilon > 0)(\exists f_\epsilon \in \mathcal{A}) f_\epsilon - \epsilon = h$$

Because every $\epsilon > 0$ can be written as $\epsilon = \alpha - \beta$ with $\alpha > 0$ and $\beta > 0$, we rewrite this equivalence as

$$\begin{aligned} h \in \mathcal{R} &\Leftrightarrow (\forall \alpha > 0)(\exists \beta > 0)(\exists f_\alpha \in \mathcal{A}) \overbrace{f_\alpha + \beta}^{\in \mathcal{N}} - \alpha = h \\ &\Leftrightarrow (\forall \alpha > 0)(\exists f_\alpha \in \mathcal{N}) f_\alpha - \alpha = h \\ &\Leftrightarrow h \in \bigcap_{\alpha>0} (\mathcal{N} - \alpha) \end{aligned}$$

From Lemma 31_{53}, Equation (1.24), $\mathrm{AD4}_{54}$ and SD4, we infer that

$$h \in \mathcal{N} \Leftrightarrow (\forall \epsilon > 0)(\exists f_\epsilon \in \mathcal{A}) f_\epsilon - \epsilon = r$$

Because every $\epsilon > 0$ can be written as $\epsilon = \alpha - \beta$ with $\alpha > 0$ and $\beta > 0$, we rewrite this equivalence as

$$\begin{aligned} h \in \mathcal{N} &\Leftrightarrow (\forall \alpha > 0)(\exists \beta > 0)(\exists f_\alpha \in \mathcal{A}) \overbrace{f_\alpha + \beta}^{\in \mathcal{N}} - \alpha = h \\ &\Leftrightarrow (\forall \alpha > 0)(\exists f_\alpha \in \mathcal{N}) f_\alpha - \alpha = h \\ &\Leftrightarrow h \in \bigcap_{\alpha>0} (\mathcal{N} - \alpha) \end{aligned}$$

□

Updating sets of strictly desirable gambles. Again we could try to use the updating rule $f \in \mathcal{N} \rfloor x_R \Leftrightarrow I_{x_R} f \in \mathcal{N}$ leading to an updated set

$$\mathrm{cylext}_{x_R}^{T} (I_{x_R} \mathcal{N} \cap \mathcal{N}).$$

The problem with this rule however is that there are situations where $I_{x_R}\mathcal{N} \cap \mathcal{N}$ is empty. However, if we know that the set of strictly desirable gambles is derived from a coherent set of acceptable gambles, then we wish the updated set $\mathcal{N}\rfloor x_R$ to be also coherent. This issue can be resolved easily: we just assume that the updated set is the vacuous model whenever $I_{x_R}\mathcal{N}\cap\mathcal{N}$ is empty.

Definition 38: *Consider a coherent set of strictly desirable gambles $\mathcal{N}\subseteq\mathcal{L}(X_N)$ and an event $X_R = x_R$ where $R\subseteq N$. Then the updated model $\mathcal{N}\rfloor x_R$ is given by*

$$\mathcal{N}\rfloor x_R := \mathcal{L}_{>0}(X_{N\setminus R})\cup\operatorname{cylext}^T_{x_R}(I_{x_R}\mathcal{N}\cap\mathcal{N}),$$

or equivalently,

$$f\in\mathcal{N}\rfloor x_R \Leftrightarrow (I_{x_R}f\in\mathcal{N})\vee(f>0).$$

The updated set of strictly desirable gambles is again a coherent set of strictly desirable gambles as the following proposition shows.

Proposition 39: *Consider a coherent set of strictly desirable gambles $\mathcal{N}\subseteq\mathcal{L}(X_N)$ and an event $X_R = x_R$ where $R\subseteq N$. Then the updated model $\mathcal{N}\rfloor x_R$ is coherent.*

Proof: SD1: Let $f\in\mathcal{L}(X_{N\setminus R})$. Because $\mathcal{N}$ avoids sure loss, there is an $x_{N\setminus R}\in\mathcal{X}_{N\setminus R}$ and $x_R\in\mathcal{X}_R$ such that $I_{x_R}(x_R)f(x_{N\setminus R})>0$ and therefore also that $f(x_{N\setminus R})>0$.

SD2: By definition, $\mathcal{L}_{>0}(X_{N\setminus R})\subseteq\mathcal{N}\rfloor x_R$.

SD3: Assume without loss of generality that $I_{x_R}\mathcal{N}\cap\mathcal{N}\neq\emptyset$. If $f\in\mathcal{N}\rfloor x_R$ and $g\in\mathcal{N}\rfloor x_R$, then $I_{x_R}f\in\mathcal{N}$ and $I_{x_R}g\in\mathcal{N}$. Because $\mathcal{N}$ is coherent and by SD3$_{57}$ it follows for $\alpha>0$ that $I_{x_R}(\alpha f+g)\in\mathcal{N}$, whence $\alpha f+g\in\mathcal{N}\rfloor x_R$.

SD4: To prove SD4$_{57}$ we may restrict ourselves to the case where $f\notin\mathcal{L}_{>0}(X_{N\setminus R})$. Under this assumption, $f\in\mathcal{N}\rfloor x_R$ implies that $I_{x_R}f\in\mathcal{N}$. We infer from Axiom SD4$_{57}$ that then there is some $\epsilon>0$ such that $I_{x_R}f-\epsilon\in\mathcal{N}$ and from Axioms SD2$_{57}$ and SD3$_{57}$ it follows that $I_{x_R}(f-\epsilon)\in\mathcal{N}$ whence

$$\mathcal{N}\rfloor x_R\subseteq\bigcup_{\epsilon>0}(\mathcal{N}\rfloor x_R+\epsilon).$$

The reverse inclusion is almost trivial: if $f\in\bigcup_{\epsilon>0}(\mathcal{N}\rfloor x_R+\epsilon)$ then there is an $\epsilon>0$ such that $f-\epsilon\in\mathcal{N}\rfloor x_R$ whence $f\in\mathcal{N}\rfloor x_R$ by SD2$_{57}$ and SD3$_{57}$. □

The following lemma will prove helpful further on.

Lemma 40: *Consider a coherent set of strictly desirable gambles* $\mathcal{N} \subseteq \mathcal{L}(X_N)$ *and an event* $X_R = x_R$ *where* $R \subseteq N$*. Then the updated set of strictly desirable gambles* $\mathcal{N}\rfloor x_R$ *is given by*

$$\mathcal{N}\rfloor x_R = \begin{cases} \text{cylext}^T_{x_R}(I_{x_R}\mathcal{N} \cap \mathcal{N}) & \text{if } I_{x_R} \in \mathcal{N}, \\ \mathcal{L}_{>0}(X_{N\setminus R}) & \text{if } I_{x_R} \notin \mathcal{N}. \end{cases}$$

or equivalently,

$$f \in \mathcal{N}\rfloor x_R \Leftrightarrow [(I_{x_R} f \in \mathcal{N}) \wedge (I_{x_R} \in \mathcal{N})] \vee (f > 0).$$

Proof: We only have to prove that $I_{x_R}\mathcal{N} \cap \mathcal{N} = \emptyset$ if and only if $I_{x_R} \notin \mathcal{N}$. Of course, $I_{x_R} \in \mathcal{N} \Rightarrow I_{x_R}\mathcal{N} \cap \mathcal{N} \neq \emptyset$ because $I_{x_R} I_{x_R} = I_{x_R}$. On the other hand, if $I_{x_R}\mathcal{N} \cap \mathcal{N} \neq \emptyset$, then by SD2$_{57}$ and SD3$_{57}$ there is some gamble $f \in \mathcal{L}_{>0}(X_{N\setminus R})$ that is not constant and with $I_{x_R} f \in \mathcal{N}$. Because $\mathcal{N}$ is coherent, there is a set of acceptable gambles $\mathcal{A}$ such that $\mathcal{N} = \bigcap_{\epsilon>0}(\mathcal{A} - \epsilon)$ whence there is an $\epsilon > 0$ such that $(I_{x_R} f - \epsilon) \in \mathcal{A}$. By A3$_{32}$ it follows that $(I_{x_R} f - \epsilon)/(\max(f) - \min(f)) \in \mathcal{A}$. Because $I_{x_R} - \epsilon/(\max(f) - \min(f)) \leq (I_{x_R} f - \epsilon)/(\max(f) - \min(f))$ it follows by A2$_{32}$ and A3$_{32}$ that $I_{x_R} - \epsilon/(\max(f) - \min(f)) \in \mathcal{A}$ and therefore $I_{x_R} \in \mathcal{N}$. □

1.4.3 Previsions

If an agent models his beliefs about the variable X through a set of acceptable gambles, then we can say which gambles he would or would not accept if they were offered to him for free. An interesting question pops up now: how much would this person—who meets a minimum of rationality requirements—be willing to pay to accept a gamble, or how much would he want to get paid to accept a gamble? The tool to get this information is what we call a **lower prevision** and we denote it by $\underline{P}_X$, or by $\underline{P}$ when it is clear from the context about which variable we are talking. Clearly, paying an amount α (positive or negative) for a gamble f will result in getting a net gamble $f - \alpha$, but this means that the question boils down to checking what is the maximal number α we can subtract from f such that $f - \alpha$ is acceptable. The cone of acceptable gambles is not necessarily closed so the maximum might not actually exist. To circumvent this problem we ask for the supremum price the agent is willing to pay.

Definition 41: *The* ***lower prevision*** *of a gamble f given a set of acceptable gambles $\mathcal{A}$ is given by*

$$\underline{P}(f) := \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{A}\}.$$

Because of the supremum in the definition, boundary information—the knowledge whether the faces of the cone of acceptable gambles are included or excluded—about the set of acceptable gambles $\mathcal{A}$ is actually ignored, which means that the definition could just as well have been formulated in terms of sets of (almost or strictly) desirable gambles.

Proposition 42: *The lower prevision of a gamble f given a set of acceptable gambles $\mathcal{A}$ is given by*

$$\begin{aligned}\underline{P}(f) &= \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{A}\}\\ &= \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{D}\}\\ &= \max\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{R}\}\\ &= \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{N}\}.\end{aligned}$$

Proof: For any $f \in \mathcal{A}$ there is a $\delta > 0$ such that for all $0 < \epsilon < \delta$:

$$f - \epsilon \in \mathcal{R} \Leftrightarrow f \in \mathcal{A} \Rightarrow f + \epsilon/2 \in \mathcal{D} \Rightarrow f + \epsilon \in \mathcal{N} \Rightarrow f + \epsilon \in \mathcal{R},$$

and consequently

$$\begin{aligned}-\epsilon + \{\alpha : f - \alpha \in \mathcal{R}\} &= \sup\{\alpha : f - \alpha - \epsilon \in \mathcal{R}\}\\ &\leq \sup\{\alpha : f - \alpha \in \mathcal{A}\}\\ &\leq \sup\{\alpha : f - \alpha + \epsilon \in \mathcal{D}\} = \epsilon/2 + \{\alpha : f - \alpha \in \mathcal{D}\}\\ &\leq \sup\{\alpha : f - \alpha + \epsilon \in \mathcal{N}\} = \epsilon + \{\alpha : f - \alpha \in \mathcal{N}\}\\ &\leq \sup\{\alpha : f - \alpha + \epsilon \in \mathcal{R}\} = \epsilon + \{\alpha : f - \alpha \in \mathcal{R}\}\end{aligned}$$

and since this holds for all $0 < \epsilon < \delta$, we see that

$$\begin{aligned}\sup\{\alpha : f - \alpha \in \mathcal{R}\} &\leq \sup\{\alpha : f - \alpha \in \mathcal{A}\} \leq \sup\{\alpha : f - \alpha \in \mathcal{D}\}\\ &\leq \sup\{\alpha : f - \alpha \in \mathcal{N}\} \leq \sup\{\alpha : f - \alpha \in \mathcal{R}\},\end{aligned}$$

so all are equal. The sup turns into max for $\mathcal{R}$ because $\mathcal{R}$ is closed by axiom AD454.

□

▷ **Example 43:** If we take the vacuous model from Example 12_{37}, then we see that

$$
\begin{aligned}
\underline{P}(f) &= \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{A}\},\\
&= \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{L}_{\geq 0}\},\\
&= \sup\{\alpha \in \mathbb{R} : f \geq \alpha\} = \min f.
\end{aligned}
$$

This example also shows why the set of indifferent gambles does not appear in Proposition 42. The only indifferent gamble for the vacuous model is the zero gamble and for most gambles $\{\alpha : f - \alpha \in \mathcal{I}\}$ is empty. One could think of using $\sup\{\alpha : f - \alpha + p \in \mathcal{I} \text{ for some } p \in \mathcal{L}_{\geq 0}\}$ which will evaluate to $\underline{P}(f)$ in case the vacuous model is assumed. In general, this equality will not hold, and in fact, it is not possible to come up with an expression for the lower prevision based on the set of indifferent gambles.

▷ **Example 44:** For the set of acceptable gambles in Example 13_{38}, the lower prevision is given by

$$
\underline{P}(f) = \sup\left\{\alpha \in \mathbb{R} : \sum_{x \in \mathcal{X}} (f(x) - \alpha) f_0(x) \geq \|f - \alpha\|_2 \|f_0\|_2 \cos(\alpha_0)\right\},
$$

and can be obtained by solving the equation

$$
\sum_{x \in \mathcal{X}} [f(x) - \underline{P}(f)] f_0(x) = \|f - \underline{P}(f)\|_2 \|f_0\|_2 \cos(\alpha_0).
$$

If we define the function $F_0 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ as[8]

$$
F_0(x, y) := \frac{f_0(x) f_0(y)}{\cos^2(\alpha_0) \|f_0\|_2^2} - \delta_{xy}, \tag{1.27}
$$

and define a, b and c as

$$
\begin{aligned}
a &:= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} F_0(x, y),\\
b &:= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} F_0(x, y) f(y),\\
c &:= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} f(x) F_0(x, y) f(y),
\end{aligned}
$$

The lower prevision $\underline{P}(f)$ is given by

$$
\underline{P}(f) = \frac{b - \sqrt{b^2 - ac}}{a} \text{ on condition that } \sum_{x \in \mathcal{X}} (f(x) - \underline{P}(f)) f_0(x) \geq 0.
$$

[8] Here δ_{xy} stands for the Kronecker delta.

In the special case where f_0 is a constant, positive gamble and we define $\overline{f} := \sum_{x\in\mathcal{X}} f(x)/|\mathcal{X}|$, we get $F_0(x,y) = \frac{1}{|\mathcal{X}|\cos^2(\alpha_0)} - \delta_{xy}$, whence

$$\frac{a}{|\mathcal{X}|} = \tan^2(\alpha_0), \frac{b}{|\mathcal{X}|} = \overline{f}\tan^2(\alpha_0) \text{ and } \frac{c}{|\mathcal{X}|} = \overline{f}^2\tan^2(\alpha_0) - \frac{\|f-\overline{f}\|_2^2}{|\mathcal{X}|},$$

and the lower prevision $\underline{\mathrm{P}}(f)$ is given by:

$$\underline{\mathrm{P}}(f) = \overline{f} - \frac{\|f-\overline{f}\|_2}{\sqrt{|\mathcal{X}|}}\cot(\alpha_0).$$

Likewise, we could think of the minimal price the agent would like to get for selling the gamble f or in other words: for what value of α does the gamble $\alpha - f$ become acceptable? This value is denoted $\overline{\mathrm{P}}(f)$ and is called the **upper prevision** of f.

Definition 45: *The **upper prevision** $\overline{\mathrm{P}}(f)$ of a gamble f is given by*

$$\begin{aligned}\overline{\mathrm{P}}(f) &:= \inf\{\alpha\in\mathbb{R}: \alpha - f \in \mathcal{A}\},\\ &= \inf\{\alpha\in\mathbb{R}: \alpha - f \in \mathcal{D}\},\\ &= \min\{\alpha\in\mathbb{R}: \alpha - f \in \mathcal{R}\},\\ &= \inf\{\alpha\in\mathbb{R}: \alpha - f \in \mathcal{N}\}.\end{aligned}$$

Observe that there is a conjugacy relation

$$\overline{\mathrm{P}}(f) = -\underline{\mathrm{P}}(-f)$$

between upper and lower previsions.

If the lower prevision is given for all gambles in a domain $\mathcal{L}(X)$ then it could also be viewed as a model describing uncertainty about X. As the boundary information of the cone of acceptable gambles is not present in the lower prevision, it is not possible to determine which set of acceptable gambles $\mathcal{A}_{\underline{\mathrm{P}}}$ is being represented by the lower prevision. This is in general not a problem, but when conditioning, the inherent ambiguity can give problems.

Definition 46: *Given a lower prevision $\underline{\mathrm{P}}$ on a domain $\mathcal{F}\subseteq\mathcal{L}$*

$$\mathcal{R}_{\underline{\mathrm{P}}} := \operatorname{regext}(\{f-\underline{\mathrm{P}}(f): f\in\mathcal{F}\}) = \bigcap_{\epsilon>0}(\mathcal{N}_{\underline{\mathrm{P}}} - \epsilon), \tag{1.28}$$

$$\mathcal{N}_{\underline{\mathrm{P}}} := \operatorname{natext}(\{f-\underline{\mathrm{P}}(f): f\in\mathcal{F}\}) = \bigcup_{\epsilon>0}(\mathcal{R}_{\underline{\mathrm{P}}} + \epsilon). \tag{1.29}$$

A lower prevision $\underline{P}$ *is said to be* ***coherent*** *if its corresponding set of almost desirable gambles* $\mathcal{R}_{\underline{P}}$ *is coherent. The* ***extension of*** $\underline{P}$ *from* $\mathcal{F}$ *to* $\mathcal{L}$ *is given by*

$$\underline{P}(f) = \sup\left\{\alpha \in \mathbb{R} \colon f - \alpha \in \mathcal{R}_{\underline{P}}\right\}. \tag{1.30}$$

We do know that

$$\mathcal{N}_{\underline{P}} \subseteq \mathcal{D}_{\underline{P}} \subset \mathcal{A}_{\underline{P}} \subseteq \mathcal{R}_{\underline{P}} \tag{1.31}$$

so we see that the represented set $\mathcal{A}_{\underline{P}}$ can be approximated by the sets $\mathcal{N}_{\underline{P}}$ and $\mathcal{R}_{\underline{P}}$. The following proposition gives an alternative expression for the sets $\mathcal{R}_{\underline{P}}$ and $\mathcal{N}_{\underline{P}}$ that is easier to interpret.

Proposition 47: *Consider a coherent lower prevision* $\underline{P}$ *on* $\mathcal{L}(\mathcal{X})$*, then the corresponding coherent set of almost desirable gambles* $\mathcal{R}_{\underline{P}}$ *and coherent set of strictly desirable gambles* $\mathcal{N}_{\underline{P}}$ *are given by:*

$$\mathcal{N}_{\underline{P}} = \{f \in \mathcal{L}(\mathcal{X}) \colon \underline{P}(f) > 0\},$$
$$\mathcal{R}_{\underline{P}} = \{f \in \mathcal{L}(\mathcal{X}) \colon \underline{P}(f) \geq 0\}.$$

Proof: We infer from Equation (1.30) that

$$\begin{aligned}\underline{P}(f) \geq 0 &\Leftrightarrow \sup\left\{\alpha \in \mathbb{R} \colon f - \alpha \in \mathcal{R}_{\underline{P}}\right\} \geq 0\\ &\Leftrightarrow (\forall \alpha \geq 0)(f + \alpha \in \mathcal{R}_{\underline{P}})\\ &\Leftrightarrow f \in \left[\bigcap_{\alpha>0} (\mathcal{R}_{\underline{P}} - \alpha) \cap \mathcal{R}_{\underline{P}}\right] \Leftrightarrow f \in \mathcal{R}_{\underline{P}},\end{aligned}$$

where the last equivalence is the result of AD4$_{54}$. In a similar way we infer that

$$\begin{aligned}\underline{P}(f) > 0 &\Leftrightarrow \sup\left\{\alpha \in \mathbb{R} \colon f - \alpha \in \mathcal{R}_{\underline{P}}\right\} > 0\\ &\Leftrightarrow (\exists \delta > 0)(\forall \epsilon > 0)(f - \delta + \epsilon \in \mathcal{R}_{\underline{P}})\\ &\Leftrightarrow (\exists \delta > 0)(f - \delta \in \mathcal{R}_{\underline{P}})\\ &\Leftrightarrow f \in \bigcup_{\delta>0} (\mathcal{R}_{\underline{P}} + \delta) \Leftrightarrow f \in \mathcal{N}_{\underline{P}},\end{aligned}$$

where Proposition 37$_{58}$ was used in the last equivalence. □

Properties of lower previsions The following properties of lower and upper previsions were proved by Walley [62].

Proposition 48: *Let* $\underline{P}$ *be defined on the linear space* $\mathcal{L}(X)$ *and assume that* f *and* g *are elements of* $\mathcal{L}(X)$*. Then the prevision* $\underline{P}$ *is coherent if and only if*

P1. $\min f \leq \underline{\mathrm{P}}(f) \leq \overline{\mathrm{P}}(f) \leq \max f$ [boundedness]

P2. $\underline{\mathrm{P}}(f) + \underline{\mathrm{P}}(g) \leq \underline{\mathrm{P}}(f+g)$ [super-additivity]

P3. $\underline{\mathrm{P}}(\alpha f) = \alpha \underline{\mathrm{P}}(f)$ *for all real* $\alpha \in \mathbb{R}_{\geq 0}$ [non-negative homogeneity]

Moreover the following properties will hold for any coherent $\underline{\mathrm{P}}$*.*

P4. $\underline{\mathrm{P}}(f+\beta) = \underline{\mathrm{P}}(f) + \beta$ *for all real* β [constant additivity]

P5. $f \leq g$ *implies that* $\underline{\mathrm{P}}(f) \leq \underline{\mathrm{P}}(g)$ [monotonicity]

P6. $\underline{\mathrm{P}}(\mu) = \mu$ *for any constant gamble* $\mu \in \mathcal{L}(X)$*.*

Proposition 49: *Given a set of acceptable gambles* $\mathcal{A}$ *with corresponding set of indifferent gambles* $\mathcal{I}$ *and lower prevision* $\underline{\mathrm{P}}$*, it holds for all* $f \in \mathcal{L}$ *and* $i \in \mathcal{I}$ *that* $\underline{\mathrm{P}}(i) = \overline{\mathrm{P}}(i) = 0$ *and* $\underline{\mathrm{P}}(f+i) = \underline{\mathrm{P}}(f)$ *and* $\overline{\mathrm{P}}(f+i) = \overline{\mathrm{P}}(f)$*.*

Proof: Because i and $-i \in \mathcal{A}$ it follows from Definition 41_{60} that $\underline{\mathrm{P}}(i) \geq 0$ and $\underline{\mathrm{P}}(-i) \geq 0$. From P6 and P2 we see that $0 \leq \underline{\mathrm{P}}(i) + \underline{\mathrm{P}}(-i) \leq \underline{\mathrm{P}}(0) = 0$ whence $\underline{\mathrm{P}}(i) = 0$. Similarly, $\underline{\mathrm{P}}(-i) = 0$ so $\overline{\mathrm{P}}(i) = 0$.

Using that $0 = \underline{\mathrm{P}}(i) = \overline{\mathrm{P}}(i)$ and super-additivity [P2] we see that $\underline{\mathrm{P}}(f) \leq \underline{\mathrm{P}}(f+i) \leq \underline{\mathrm{P}}(f) + \overline{\mathrm{P}}(i) = \underline{\mathrm{P}}(f)$. The last inequality comes from $\underline{\mathrm{P}}(f+i) + \underline{\mathrm{P}}(-i) \leq \underline{\mathrm{P}}(f)$ together with the conjugacy property of the lower and upper prevision. □

If $\overline{\mathrm{P}}(f) = \underline{\mathrm{P}}(f)$ for any gamble f on X, then we say that P defined by $\underline{\mathrm{P}}(f) = \overline{\mathrm{P}}(f) = \mathrm{P}(f)$ is a **linear prevision**. It follows immediately from the conjugacy relation, P2 and P3 that a linear prevision P is indeed a linear functional. It is interesting to lay bare the link between linear previsions and resolved models.

Proposition 50: *With a resolved model* $\mathcal{M} \in \mathbb{M}$ *there corresponds a linear prevision* $\mathrm{P}_{\mathcal{M}}$*. Models that correspond to a linear prevision* P *are resolved.*

Proof: The set $MP(f) := \{\alpha : f - \alpha \in \mathcal{M}\}$ is a downset of $\mathbb{R}$:

$$\alpha_1 \leq \alpha_2 \Rightarrow f - \alpha_1 \leq f - \alpha_2 \Rightarrow [\alpha_2 \in MP(f) \Rightarrow \alpha_1 \in MP(f)],$$

so $MP(f)$ has the form $\left(-\infty, \underline{\mathrm{P}}_{\mathcal{M}}(f)\right)$ or $(-\infty, \underline{\mathrm{P}}_{\mathcal{M}}(f)]$. In any case, we see that

$$\underline{\mathrm{P}}_{\mathcal{M}}(f) = \sup\{\alpha : f - \alpha \in \mathcal{M}\} = \inf\{\beta : f - \beta \notin \mathcal{M}\}$$

and since $\mathcal{M}$ is resolved $f - \beta \notin \mathcal{M} \Rightarrow \beta - f \in \mathcal{M}$, so

$$\underline{\mathrm{P}}_{\mathcal{M}}(f) = \inf\{\beta : \beta - f \in \mathcal{M}\} = \overline{\mathrm{P}}_{\mathcal{M}}(f).$$

To prove that a model can correspond with a linear prevision only if it is resolved, we infer from Definition 46[63] and Proposition 47[64] that

$$f \notin \mathcal{R}_{\mathrm{P}} \Rightarrow \mathrm{P}(f) < 0 \Leftrightarrow \mathrm{P}(-f) > 0 \Leftrightarrow -f \in \mathcal{R}_{\mathrm{P}}.$$

But then it follows from Proposition 15[39] that $\mathcal{R}_{\mathrm{P}}$ must be resolved. □

Conditional previsions The conditional lower prevision $\underline{\mathrm{P}}(\cdot|x_r)$ on the linear space $\mathcal{L}(X_{N\setminus R})$ for a conditioning event $X_R = x_r$ can be calculated from its corresponding updated set of acceptable gambles.

Definition 51: *Consider a coherent set of acceptable gambles $\mathcal{A}$ on $\mathcal{L}(X_N)$, a conditioning event $X_R = x_R$ ($R \subseteq N$) and a gamble $f \in \mathcal{L}(X_{N\setminus R})$. Then the lower prevision and upper prevision of f conditional on $X_R = x_R$ are defined as:*

$$\underline{\mathrm{P}}(f|x_R) = \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{A}\rfloor x_R\},$$
$$\overline{\mathrm{P}}(f|x_R) = \inf\{\alpha \in \mathbb{R} : \alpha - f \in \mathcal{A}\rfloor x_R\}.$$

What happens if $\underline{\mathrm{P}}$ is given instead of a coherent set of acceptable gambles? We know that $\mathcal{A}_{\underline{\mathrm{P}}}$ is not uniquely determined by $\underline{\mathrm{P}}$, and so it cannot be expected that $\underline{\mathrm{P}}(\cdot|\cdot)$ is uniquely defined. We know that we can bound the compatible sets of acceptable gambles, and from Inequality (1.31) we know that $\mathcal{N}_{\underline{\mathrm{P}}} \subseteq \mathcal{A}_{\underline{\mathrm{P}}} \subseteq \mathcal{R}_{\underline{\mathrm{P}}}$ and consequently, we are able to confine the lower prevision. The conditional previsions, corresponding to these extremes are called the lower prevision under natural extension $\underline{\mathrm{N}}$ if $\mathcal{N}$ is assumed and the lower prevision under regular extension $\underline{\mathrm{R}}$ if $\mathcal{R}$ is assumed.

Definition 52: *Given a coherent lower prevision $\underline{\mathrm{P}}$ on $\mathcal{L}(X_N)$ and a conditioning event $X_R = x_R$ then the conditional lower prevision under regular and natural extension are respectively given by:*

$$\underline{\mathrm{R}}(f|x_R) := \sup\left\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{R}_{\underline{\mathrm{P}}}\Big\rfloor x_R\right\} \tag{1.32}$$
$$\underline{\mathrm{N}}(f|x_R) := \sup\left\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{N}_{\underline{\mathrm{P}}}\Big\rfloor x_R\right\} \tag{1.33}$$

with $\mathcal{R}_{\underline{\mathrm{P}}}$ and $\mathcal{N}_{\underline{\mathrm{P}}}$ as in Definition 46[63].

Although the exact lower prevision cannot be known from the information available, we can safely assume that it resides somewhere in the interval $[\underline{\mathrm{N}}, \underline{\mathrm{R}}]$. In practical applications, usually one of the two bounds is assumed, either natural or regular extension. It is interesting to note that $\mathcal{N}_{\underline{\mathrm{P}}}\rfloor x_R$ will differ from $\mathcal{R}_{\underline{\mathrm{P}}}\rfloor x_R$ if and only if $I_{x_R}\mathcal{N}_{\underline{\mathrm{P}}} \cap \mathcal{N}_{\underline{\mathrm{P}}} = \emptyset$, as the following proposition indicates.

Proposition 53 (Generalised Bayes Rule): *The lower prevision of a gamble $f \in \mathcal{L}(x_{N\setminus R})$, conditional on an event $X_R = x_R$ with $R \subseteq N$ for a given coherent (unconditional) lower prevision $\underline{\mathrm{P}}$ on $\mathcal{L}(X_N)$ is the solution of*

$$\begin{cases} \underline{\mathrm{P}}(I_{x_R}[f - \underline{\mathrm{P}}(f|x_R)]) = 0 & \text{if } \underline{\mathrm{P}}(I_{x_R}) > 0, \\ \underline{\mathrm{P}}(f|x_R) = \min f & \text{if } \overline{\mathrm{P}}(I_{x_R}) = 0. \end{cases} \tag{1.34}$$

When $\underline{\mathrm{P}}(I_{x_R}) = 0 < \overline{\mathrm{P}}(I_{x_R})$, then the conditional lower prevision $\underline{\mathrm{P}}(\cdot|x_R)$ is not uniquely determined , but has bounds

$$\begin{aligned} \underline{\mathrm{N}}(f|x_R) &= \min f, \\ \underline{\mathrm{R}}(f|x_R) &= \max\{\alpha \in \mathbb{R}: \underline{\mathrm{P}}(I_{x_R}[f-\alpha]) \geq 0\}. \end{aligned}$$

Proof: It follows from P1$_{65}$ and Proposition 47$_{64}$ that $I_{x_R} \notin \mathcal{N}_{\underline{\mathrm{P}}} \Leftrightarrow \underline{\mathrm{P}}(I_{x_R}) = 0$. From Lemma 40$_{60}$ we infer then that for any $f \in \mathcal{L}(X_{N\setminus R})$

$$\underline{\mathrm{N}}(f|x_R) = \begin{cases} \sup\left\{\alpha: I_{x_R}(f-\alpha) \in \mathcal{N}_{\underline{\mathrm{P}}}\right\} & \text{if } \underline{\mathrm{P}}(I_{x_R}) > 0, \\ \min f & \text{if } \underline{\mathrm{P}}(I_{x_R}) = 0. \end{cases}$$

Analogously, it is not too difficult to show that $-I_{x_r} \in \mathcal{R}_{\underline{\mathrm{P}}} \Leftrightarrow \overline{\mathrm{P}}(I_{x_R}) = 0$. From Definition 38$_{59}$ we infer that

$$\underline{\mathrm{R}}(f|x_R) = \begin{cases} \max\left\{\alpha: I_{x_R}(f-\alpha) \in \mathcal{R}_{\underline{\mathrm{P}}}\right\} & \text{if } \overline{\mathrm{P}}(I_{x_R}) > 0, \\ \min f & \text{if } \overline{\mathrm{P}}(I_{x_R}) = 0. \end{cases}$$

We see that $\mathcal{N}_{\underline{\mathrm{P}}}\rfloor x_R$ and $\mathcal{R}_{\underline{\mathrm{P}}}\rfloor x_R$ will result in a different lower prevision only if $\underline{\mathrm{P}}(I_{x_R}) = 0$. If $\overline{\mathrm{P}}(I_{x_R}) = 0$, then both bounds coincide with the vacuous model. Therefore, assuming that $0 < \underline{\mathrm{P}}(I_{x_R}) \leq \overline{\mathrm{P}}(I_{x_R})$, we see that the value of $\underline{\mathrm{P}}(f|x_R)$ is uniquely determined and given by

$$\underline{\mathrm{P}}(f|x_R) = \max\left\{\alpha \in \mathbb{R}: I_{x_R}[f-\alpha] \in \mathcal{R}_{\underline{\mathrm{P}}}\right\}.$$

Observing that $I_{x_R}[f-\alpha] \in \mathcal{R}_{\underline{\mathrm{P}}} \Leftrightarrow \underline{\mathrm{P}}(I_{x_R}[f-\alpha]) \geq 0$, we can rewrite the conditional lower prevision in terms of the unconditional one

$$\underline{\mathrm{P}}(f|x_R) = \max\{\alpha \in \mathbb{R}: \underline{\mathrm{P}}(I_{x_R}[f-\alpha]) \geq 0\}.$$

From Lemma 54 we conclude that if $\underline{\mathrm{P}}(I_{x_R}) > 0$ then there is a unique root for $\underline{\mathrm{P}}(I_{x_R}[f-\alpha])$ whence

$$\underline{\mathrm{P}}(I_{x_R}[f - \underline{\mathrm{P}}(f|x_R)]) = 0. \qquad \square$$

Equation (1.34) is called the **generalised Bayes Rule** [62, par. 6.4.1].

Lemma 54: *Let* $\underline{\mathrm{P}}$ *be a coherent prevision. Then the map*

$$\rho_f : \mathbb{R} \to \mathbb{R} : \alpha \to \underline{\mathrm{P}}\left(I_{x_R}[f-\alpha]\right)$$

is continuous, concave and non-increasing. ρ_f *is decreasing when* $\underline{\mathrm{P}}\left(I_{x_R}\right) > 0$.

Proof: Assume that $\alpha_1 \le \alpha_2$, then by coherence of $\underline{\mathrm{P}}$ we infer from Proposition 48_{64} that

$$\underline{\mathrm{P}}(I_{x_R}[f-\alpha_1]) - \underline{\mathrm{P}}(I_{x_R}[f-\alpha_2]) \le \overline{\mathrm{P}}(I_{x_R}[\alpha_2-\alpha_1]) = (\alpha_2-\alpha_1)\overline{\mathrm{P}}(I_{x_R}),$$

and we conclude that $\left|\rho_f(\alpha_1) - \rho_f(\alpha_2)\right| \le |\alpha_1-\alpha_2|\overline{\mathrm{P}}(I_{x_R})$ for all $\alpha_1, \alpha_2 \in \mathbb{R}$, which implies that ρ_f is (Lipschitz) continuous.

To show that ρ_f is concave, we just apply Properties $\mathrm{P2}_{65}$ and $\mathrm{P3}_{65}$:

$$\begin{aligned}\rho_f(\epsilon\alpha_1 + (1-\epsilon)\alpha_2) &= \underline{\mathrm{P}}(I_{x_R}[\epsilon f - \epsilon\alpha_1 + (1-\epsilon)f - (1-\epsilon)\alpha_2]) \\ &\ge \epsilon\underline{\mathrm{P}}(I_{x_R}[f-\alpha_1]) + (1-\epsilon)\underline{\mathrm{P}}(I_{x_R}[f-\alpha_2]) \\ &= \epsilon\rho_f(\alpha_1) + (1-\epsilon)\rho_f(\alpha_2).\end{aligned}$$

From property $\mathrm{P5}_{65}$, it follows immediately that ρ_f is non-increasing.

To prove that ρ_f is decreasing when $\underline{\mathrm{P}}(I_{x_R}) > 0$ we let $\alpha \in \mathbb{R}$ and $\epsilon \in \mathbb{R}_{>0}$ and infer from $\mathrm{P2}_{65}$ that

$$\rho_f(\alpha - \epsilon) \ge \rho(\alpha) + \underline{\mathrm{P}}(\epsilon I_{x_R}).$$

By using $\mathrm{P3}_{65}$ and $\underline{\mathrm{P}}(I_{x_R}) > 0$ we infer that

$$\rho_f(\alpha - \epsilon) > \rho(\alpha),$$

whence ρ_f must be a decreasing function. □

1.5 Structural judgements

1.5.1 Irrelevance and independence

A coherent set $\mathrm{marg}_Y(\mathcal{A}\rfloor X = x)$ represents your current attitude towards acceptance of gambles in $\mathcal{L}(Y)$ when X assumes a value in $x \in \mathcal{X}$.[9] If your attitude towards accepting gambles in $\mathcal{L}(Y)$ does not depend on whether

[9] Following de Cooman, Miranda & Zaffalon [17] we demand only the independence towards atoms instead of independence towards every possible subset.

you have learned something about the possible values that X can assume, then we say that X is epistemically irrelevant to Y. It means that it does not matter whether we do or do not learn something about X when having to make decisions about Y.

Definition 55: *Let $\mathcal{A} \in \mathcal{L}(X_N)$ be a coherent set of acceptable gambles modelling your beliefs on X_N, and let R and V be any disjoint subsets of N. Then we say that X_R is* ***epistemically irrelevant*** *to X_V and denote this $X_R \mathrel{\text{EI}} X_V$ if learning the outcome of X_R will not alter your beliefs about X_V:*

$$X_R \mathrel{\text{EI}} X_V \Leftrightarrow (\forall x_R \in \mathcal{X}_R)\left(\mathrm{marg}_V(\mathcal{A}\rfloor x_R) = \mathrm{marg}_V(\mathcal{A})\right). \tag{1.35}$$

Very often, an irrelevance statement is made and some marginals are given and the question is which joint model coincides with these assessments and if possible, which compatible model is the least committal one.

Theorem 56 (irrelevant natural extension): *The least committal, jointly coherent set of acceptable gambles on $\mathcal{L}(X,Y)$ that expresses the epistemic irrelevance $X \mathrel{\text{EI}} Y$ given the separately coherent marginals $\mathcal{A}_X$ and $\mathcal{A}_Y$ is given by*

$$\begin{aligned}\mathcal{A}_{X \mathrel{\text{EI}} Y} &:= \sum_{x\in\mathcal{X}} I_{\{x\}}\mathcal{A}_Y + \mathrm{proj}_X^T(\mathcal{A}_X)\\ &:= \sum_{x\in\mathcal{X}} I_{\{x\}}\mathcal{A}_Y + \mathcal{A}_X.\end{aligned}$$

Proof: It is clear that $\mathcal{A}_{X \mathrel{\text{EI}} Y}$ should at least include $\sum_{x\in\mathcal{X}} I_{\{x\}}\mathcal{A}_Y + \mathcal{A}_X$.

The rest of the proof follows from the fact that the irrelevant natural extension is a special type of marginal extension (see Theorem 30_{51}) where all conditional models are given by $\mathcal{A}_Y\rfloor x = \mathcal{A}_Y$ and the marginal model is $\mathcal{A}_X$. □

▷ **Example 57:** Consider two variables X and Y with $\mathcal{X} = \{x_1, x_2\}$ and $\mathcal{Y} = \{y_1, y_2\}$. The set of acceptable gambles $\mathcal{A}_X$ is the nonnegative linear hull of the two extreme gambles $a := -I_{\{x_1\}} + 2I_{\{x_2\}}$ and $b := 3I_{\{x_1\}} - I_{\{x_2\}}$. $\mathcal{A}_Y$ is the nonnegative linear hull of the two extreme gambles $c := -I_{\{y_1\}} + 4I_{\{y_2\}}$ and $d := I_{\{y_1\}} - I_{\{y_2\}}$.

$$\mathcal{A}_X = \mathrm{posi}\{a, b, 0\} = \{\alpha a + \beta b \colon \alpha, \beta \geq 0\}, \tag{1.36}$$

$$\mathcal{A}_Y = \mathrm{posi}\{c, d, 0\} = \{\gamma c + \delta d \colon \gamma, \delta \geq 0\}. \tag{1.37}$$

If we assume that X is epistemically irrelevant to Y: $X \mathrel{\text{EI}} Y$, then we use Theorem 56 to get

$$\mathcal{A}_{X \mathrel{\text{EI}} Y} := \sum_{x\in\mathcal{X}} I_{\{x\}}\mathcal{A}_Y + \mathrm{proj}_X^T(\mathcal{A}_X),$$

$$
\begin{aligned}
&= I_{\{x_1\}}\mathcal{A}_Y + I_{\{x_2\}}\mathcal{A}_Y + \operatorname{proj}_X^T(\mathcal{A}_X),\\
&= \left\{ I_{\{x_1\}}[f(x_1)+g_1] + I_{\{x_2\}}[f(x_2)+g_2] : f \in \mathcal{A}_X \wedge g_1, g_2 \in \mathcal{A}_Y \right\}.
\end{aligned} \tag{1.38}
$$

If we define a basis $\mathcal{B} = \left\{ I_{\{x_1,y_1\}}, I_{\{x_1,y_2\}}, I_{\{x_2,y_1\}}, I_{\{x_2,y_2\}} \right\}$, then the vector representation $[\mathcal{A}_{X \,\mathsf{EI}\, Y}]_{\mathcal{B}}$ is given by

$$
[\mathcal{A}_{X \,\mathsf{EI}\, Y}]_{\mathcal{B}} = \left\{ \begin{pmatrix} -\alpha + 3\beta - \gamma_1 + \delta_1 \\ -\alpha + 3\beta + 4\gamma_1 - \delta_1 \\ 2\alpha - \beta - \gamma_2 + \delta_2 \\ 2\alpha - \beta + 4\gamma_2 - \delta_2 \end{pmatrix} : \alpha, \beta, \gamma_1, \gamma_2, \delta_1, \delta_2 \geq 0 \right\}.
$$

The fact that the Y-marginal is equal to $\mathcal{A}_Y$ can be seen the easiest from Equation (1.38).

$$
\operatorname{marg}_{\mathcal{Y}}(\mathcal{A}_{X \,\mathsf{EI}\, Y}) = \{ f(x_1) + g_1 : f \in \mathcal{A}_X, g_1, g_2 \in \mathcal{A}_Y, f(x_1) + g_1 = f(x_2) + g_2 \}.
$$

Note that $g_1 = g_2 + c$ where c is a constant. Because $\mathcal{A}_X$ is coherent we know that $\max f \geq 0$ and therefore either $f(x_2) + g_2 \in \mathcal{A}_Y$ or $f(x_1) + g_1 \in \mathcal{A}_Y$. Because $f(x_1) + g_1 = f(x_2) + g_2$ we see that $f(x_1) + g_1 \in \mathcal{A}_Y$.

The X-marginal $\operatorname{marg}_{\mathcal{X}}(\mathcal{A}_{X \,\mathsf{EI}\, Y})$ is given by

$$
\left\{ I_{\{x_1\}}[f(x_1)+g_1] + I_{\{x_2\}}[f(x_2)+g_2] : f \in \mathcal{A}_X, g_1, g_2 \in \mathcal{A}_Y \cap \mathcal{L}(X) \right\}.
$$

Because $\mathcal{A}_Y$ is coherent, $\mathcal{A}_Y \cap \mathcal{L}(X)$ consists solely of non-negative constant gambles (including zero) whence $\operatorname{marg}_{\mathcal{X}}(\mathcal{A}_{X \,\mathsf{EI}\, Y}) = \mathcal{A}_X$.

To end this example, we check whether $\operatorname{marg}_{\mathcal{Y}}(\mathcal{A}_{X \,\mathsf{EI}\, Y}|x_1) = \mathcal{A}_Y$. Again, from Equation (1.38) we get that

$$
\operatorname{marg}_{\mathcal{Y}}(\mathcal{A}_{X \,\mathsf{EI}\, Y}|x_1) = \{ f(x_1) + g : f \in \mathcal{A}_X, g, h \in \mathcal{A}_Y, f(x_2) + h = 0 \}.
$$

Because $f(x_2) + h$ has to be equal to 0, h is a constant gamble which is non-negative because of coherence of $\mathcal{A}_Y$. But this means that $f(x_2) \leq 0$ and because of coherence of $\mathcal{A}_X$, $f(x_1)$ has to be non-negative. Thus $\operatorname{marg}_{\mathcal{Y}}(\mathcal{A}_{X \,\mathsf{EI}\, Y}|x_1) = \mathcal{A}_Y$.

If learning about X will not alter our beliefs about Y and vice versa, learning about Y does not alter our beliefs about X then we say that X and Y are **epistemically independent**.

Definition 58: *Two variables X and Y are said to be* ***epistemically independent*** *if X is epistemically irrelevant to Y and Y is epistemically irrelevant to X*

$$
X \,\mathsf{ind}\, Y \Leftrightarrow X \,\mathsf{EI}\, Y \text{ and } Y \,\mathsf{EI}\, X.
$$

As in Theorem 56[69] we can try to extend marginals again, but now under the independence criterion.

Lemma 59 (independent natural extension): *The least committal set on $\mathcal{L}(X,Y)$ embedding epistemic independence between the two variables X and Y given the separately coherent X-marginal $\mathcal{A}_X$ and Y-marginal $\mathcal{A}_Y$ is given by*

$$\mathcal{A}_X \otimes \mathcal{A}_Y := \sum_{x \in \mathcal{X}} I_{\{x\}} \mathcal{A}_Y + \sum_{y \in \mathcal{Y}} I_{\{y\}} \mathcal{A}_X,$$

This extension is jointly coherent with the local models.

Proof: Clearly $\mathcal{A}_X \otimes \mathcal{A}_Y$ should include both $\mathcal{A}_{X\,\mathsf{EI}\,Y}$ and $\mathcal{A}_{Y\,\mathsf{EI}\,X}$ and $\operatorname{ext}(\mathcal{A}_{X\,\mathsf{EI}\,Y} \cup \mathcal{A}_{Y\,\mathsf{EI}\,X})$ is the smallest—possibly coherent—set doing this. This natural extension will be coherent if and only if $\mathcal{A}_{X\,\mathsf{EI}\,Y} \cup \mathcal{A}_{Y\,\mathsf{EI}\,X}$ avoids sure loss. Given the coherence of the epistemic irrelevant extensions, $\operatorname{ext}(\mathcal{A}_{X\,\mathsf{EI}\,Y} \cup \mathcal{A}_{Y\,\mathsf{EI}\,X})$ is equal to $\sum_{x \in \mathcal{X}} I_{\{x\}} \mathcal{A}_Y + \mathcal{A}_X + \sum_{y \in \mathcal{Y}} I_{\{y\}} \mathcal{A}_X + \mathcal{A}_Y$ which can also be written as

$$\sum_{x \in \mathcal{X}} I_{\{x\}} \mathcal{A}_Y + \sum_{y \in \mathcal{Y}} I_{\{y\}} \mathcal{A}_X.$$

Next we show that $\operatorname{marg}_{\mathcal{X}}(\mathcal{A}_X \otimes \mathcal{A}_Y) = \mathcal{A}_X$. Let $f \in ((\mathcal{A}_X \otimes \mathcal{A}_Y) \cap \mathcal{L}(X))$. This means that there are $h_x \in \mathcal{A}_Y$ and $h_y \in \mathcal{A}_X$ such that

$$f = f(\cdot, y) = \sum_{x \in \mathcal{X}} I_{\{x\}} h_x(y) + h_y \quad \text{for any } y \in \mathcal{Y}.$$

But this is an element of $\mathcal{A}_{X\,\mathsf{EI}\,Y} \cap \mathcal{L}(X) = \operatorname{marg}_{\mathcal{X}}(\mathcal{A}_{X\,\mathsf{EI}\,Y}) = \mathcal{A}_X$ and therefore $\operatorname{marg}_{\mathcal{X}}(\mathcal{A}_X \otimes \mathcal{A}_Y) \subseteq \mathcal{A}_X$. Because $\mathcal{A}_X \subseteq \operatorname{marg}_{\mathcal{X}}(\mathcal{A}_X \otimes \mathcal{A}_Y)$ by definition, we have shown that $\operatorname{marg}_{\mathcal{X}}(\mathcal{A}_X \otimes \mathcal{A}_Y) = \mathcal{A}_X$. This also shows that $\mathcal{A}_X \otimes \mathcal{A}_Y$ avoids sure loss because if it did not then $\mathcal{A}_X \otimes \mathcal{A}_Y \cap \mathcal{L}(X) = \mathcal{L}(X) \neq \mathcal{A}_X$.

We know from Definition 58 that $\operatorname{marg}_{\mathcal{X}}((\mathcal{A}_X \otimes \mathcal{A}_Y)|y) \supseteq \mathcal{A}_X$ so we only have to show the reverse in the remainder of the proof. We start from the expression $\operatorname{marg}_{\mathcal{X}}((\mathcal{A}_X \otimes \mathcal{A}_Y)|y) = \mathcal{L}(X) \cap I_{\{y\}}(\mathcal{A}_X \otimes \mathcal{A}_Y) \cap (\mathcal{A}_X \otimes \mathcal{A}_Y)$. Because we know that $\operatorname{marg}_{\mathcal{X}}(\mathcal{A}_X \otimes \mathcal{A}_Y) = \mathcal{A}_X$, we see that $\operatorname{marg}_{\mathcal{X}}((\mathcal{A}_X \otimes \mathcal{A}_Y)|y) = I_{\{y\}}(\mathcal{A}_X \otimes \mathcal{A}_Y) \cap \mathcal{A}_X \subseteq \mathcal{A}_X$. From symmetry it follows that $\operatorname{marg}_{\mathcal{Y}}((\mathcal{A}_X \otimes \mathcal{A}_Y)|x) = \mathcal{A}_Y$. □

Independence can again be used when the marginals for a multitude of variables X_N are given. The product $\otimes_{n \in N} \mathcal{A}_{X_n}$ can be constructed in a recursive manner because the operator $\otimes$ is commutative and associative. A product that is built this way is **completely independent**, meaning that for any subsets S, R of N with $S \cap R = \emptyset$, $X_R \otimes X_S$. In general, pairwise and complete independence are not the same however.

Proposition 60: *The operator* $\otimes$ *is commutative and associative. The independent product of a finite number of random variables* X_N, $N := \{1, 2, \dots, n\}$ *with corresponding coherent sets of acceptable gambles* $\mathcal{A}_{X_i}$, $i \in N$ *is therefore given by*

$$\otimes_{i\in N}\mathcal{A}_{X_i} := \sum_{i\in N}\sum_{z\in X_{N\setminus\{i\}}} I_{\{z\}}\mathcal{A}_{X_i}.$$

Proof: *It follows at once from Lemma 59 that* $\otimes$ *is commutative. The associativity of* $\otimes$ *follows from*

$$\begin{aligned}
&(\mathcal{A}_{X_1}\otimes\mathcal{A}_{X_2})\otimes\mathcal{A}_{X_3}\\
&\quad=\left(\sum_{x_2\in\mathcal{X}_2} I_{\{x_2\}}\mathcal{A}_{X_1}+\sum_{x_1\in\mathcal{X}_1} I_{\{x_1\}}\mathcal{A}_{X_2}\right)\otimes\mathcal{A}_{X_3}\\
&\quad=\sum_{x_3\in\mathcal{X}_3} I_{\{x_3\}}\left(\sum_{x_2\in\mathcal{X}_2} I_{\{x_2\}}\mathcal{A}_{X_1}+\sum_{x_1\in\mathcal{X}_1} I_{\{x_1\}}\mathcal{A}_{X_2}\right)+\sum_{z_3\in\mathcal{X}_{1:2}} I_{\{z_3\}}\mathcal{A}_{X_3}\\
&\quad=\sum_{z_1\in\mathcal{X}_{2:3}} I_{\{z_1\}}\mathcal{A}_{X_1}+\sum_{z_2\in\mathcal{X}_{1,2}} I_{\{z_2\}}\mathcal{A}_{X_2}+\sum_{z_3\in\mathcal{X}_{1:2}} I_{\{z_3\}}\mathcal{A}_{X_3}\\
&\quad=\sum_{z_1\in\mathcal{X}_{2:3}} I_{\{z_1\}}\mathcal{A}_{X_1}+\sum_{x_1\in\mathcal{X}_1} I_{\{x_1\}}\left(\sum_{x_3\in\mathcal{X}_3} I_{\{x_3\}}\mathcal{A}_{X_2}+\sum_{x_2\in\mathcal{X}_2} I_{\{x_2\}}\mathcal{A}_{X_3}\right)\\
&\quad=\mathcal{A}_{X_1}\otimes\left(\mathcal{A}_{X_2}\otimes\mathcal{A}_{X_3}\right).
\end{aligned}$$

To see that the necessary irrelevancies hold, observe for any $R, S \subseteq N$ *with* $R\cap S=\emptyset$, *that*

$$\begin{aligned}
\operatorname{marg}_R\left(\otimes_{n\in N}\mathcal{A}_{X_n}\middle\rfloor x_S\right)&=\operatorname{marg}_R\left(\left[(\otimes_{r\in R}\mathcal{A}_{X_r})\otimes(\otimes_{s\in S}\mathcal{A}_{X_s})\right]\middle\rfloor x_S\right)\\
&=\otimes_{r\in R}\mathcal{A}_{X_r},
\end{aligned}$$

where the last equality is a consequence of Lemma 59. □

As De Cooman and Miranda [17] prove, independent sets have an interesting factorisation property.

Proposition 61 (Factorisation): *Let* $\mathcal{A}_N := \otimes_{n\in N}\mathcal{A}_{X_n} \subseteq \mathcal{L}(X_N)$ *be the independent product of the coherent sets of acceptable gambles* $\mathcal{A}_n$. *Then for all disjoint subsets* I *and* O *of* N *and for all* $f \in \mathcal{L}(X_O)$:

$$f \in \mathcal{A}_N \Leftrightarrow (\forall g \in \mathcal{L}_{\geq 0}(X_I))\, fg \in \mathcal{A}_N. \tag{1.39}$$

Proof: Fix arbitrary disjoint subsets I and O of N and any $f \in \mathcal{L}(X_O)$; We show that Equation (1.39) holds. The '$\Leftarrow$' part is trivial, For the '$\Rightarrow$' part, assume that $f \in \mathcal{A}_N$ and consider any $g \in \mathcal{L}_{\geq 0}(X_I)$. We have to show that $fg \in \mathcal{A}_N$. Since $g = \sum_{x_I \in \mathcal{X}_I} I_{\{x_I\}} g(x_I)$, we see that $fg = \sum_{x_I \in \mathcal{X}_I} g(x_I) I_{\{x_I\}} f$. Now since $f \in \operatorname{marg}_O(\mathcal{A}_N)$, we infer from the independence of $\mathcal{A}_N$ that $f \in \mathcal{A}_N \rfloor x_I$ and therefore $I_{\{x_I\}} f \in \mathcal{A}_N$ for all $x_I \in \mathcal{X}_I$. We conclude that fg is a positive linear combination of elements $I_{\{x_I\}} f$ of $\mathcal{A}_N$, and therefore belongs to $\mathcal{A}_N$ by coherence. □

The independent natural product of a finite collection of marginal lower previsions $\underline{P}_n$, $n \in N$ is defined by

$$(\otimes_{n \in N} \underline{P}_n)(f) := \sup\left\{\alpha \in \mathbb{R} : f - \alpha \in \otimes_{n \in N} \mathcal{N}_{\underline{P}_n}\right\}.$$

We used sets of strictly desirable gambles in the definition, but any set of acceptable gambles that is compatible with the marginal lower prevision, would result in the same lower prevision. Moreover, some of the interesting properties of the $\otimes$ operator are inherited by this product.

Proposition 62: *Let $f \in \mathcal{L}(\mathcal{X}_N)$. The independent natural product of a collection of coherent marginal lower previsions $\underline{P}_n$, $n \in N$ is coherent and satisfies*

$$(\otimes_{n \in N} \underline{P}_n)(f) = \max\left\{\alpha \in \mathbb{R} : f - \alpha \in \otimes_{n \in N} \mathcal{R}_{\underline{P}_n}\right\}.$$

If $h \in \mathcal{L}(\mathcal{X}_R)$ with $R \subseteq N$, then

$$(\otimes_{n \in N} \underline{P}_n)(h) = (\otimes_{r \in R} \underline{P}_r)(h). \tag{1.40}$$

Proof: If we define the closure $\operatorname{cl}(\mathcal{A})$ of a set $\mathcal{A}$ by $\operatorname{cl}(\mathcal{A}) := \bigcap_{\epsilon > 0}(\mathcal{A} - \epsilon)$, then we infer from Proposition 42$_{61}$, that

$$\begin{aligned}(\otimes_{n \in N} \underline{P}_n)(f) &= \max\left\{\alpha \in \mathbb{R} : f - \alpha \in \operatorname{cl}\left(\otimes_{n \in N} \mathcal{N}_{\underline{P}_n}\right)\right\} \\ &= \max\left\{\alpha \in \mathbb{R} : f - \alpha \in \otimes_{n \in N} \operatorname{cl}\left(\mathcal{N}_{\underline{P}_n}\right)\right\} \\ &= \max\left\{\alpha \in \mathbb{R} : f - \alpha \in \otimes_{n \in N} \mathcal{R}_{\underline{P}_n}\right\}.\end{aligned}$$

Here we used Lemma 63 in the middle, and Definition 46$_{63}$ in the last step.

The coherence of $\otimes_{n \in N} \underline{P}_n$ is (by Definition 46$_{63}$) an immediate consequence of the coherence of $\otimes_{n \in N} \mathcal{R}_{\underline{P}_n}$ (by Proposition 60).

To prove Equation (1.40), it is sufficient to notice that

$$\begin{aligned}\operatorname{proj}_R^T(h) - \alpha \in \otimes_{n \in N} \mathcal{R}_{\underline{P}_n} &\Leftrightarrow \operatorname{proj}_R^T(h - \alpha) \in \otimes_{n \in N} \mathcal{R}_{\underline{P}_n} \\ &\Leftrightarrow (h - \alpha) \in \operatorname{marg}_R\left(\otimes_{n \in N} \mathcal{R}_{\underline{P}_n}\right)\end{aligned}$$

$$\Leftrightarrow (h-\alpha) \in \otimes_{r\in R}\mathcal{R}_{\underline{P}_r},$$

where the last equivalence follows by Proposition 6072. □

Lemma 63: *Let $\mathcal{A}_n \in \mathcal{L}(X_n)$, $n \in N$, be coherent sets of acceptable gambles and define* $\mathrm{cl}(\mathcal{A}) := \bigcap_{\epsilon>0}(\mathcal{A}-\epsilon)$*. Then*

$$\mathrm{cl}\,(\otimes_{n\in N}\mathcal{A}_n) = \otimes_{n\in N}\,(\mathrm{cl}\,\mathcal{A}_n),$$

Proof: It follows from the definition of cl, A232 and A332 that

$$\begin{aligned} f \in \mathrm{cl}\,(\otimes_{n\in N}\mathcal{A}_n) &\Leftrightarrow (\forall\delta>0) f+\delta \in \otimes_{n\in N}\mathcal{A}_n \\ &\Rightarrow (\forall\delta>0)(\forall\epsilon_{n,z}>0) f+\delta+\sum_{n\in N}\sum_{z\in X_{N\setminus\{n\}}} I_{\{z\}}\epsilon_{n,z} \in \otimes_{n\in N}\mathcal{A}_n \\ &\Rightarrow (\forall\delta>0)(\forall\epsilon_{n,z}>0) f+\sum_{n\in N}\sum_{z\in X_{N\setminus\{n\}}} I_{\{z\}}(\epsilon_{n,z}+\delta/N) \in \otimes_{n\in N}\mathcal{A}_n \\ &\Rightarrow (\forall\epsilon_{n,z}>0) f+\sum_{n\in N}\sum_{z\in X_{N\setminus\{n\}}} I_{\{z\}}\epsilon_{n,z} \in \otimes_{n\in N}\mathcal{A}_n \\ &\Rightarrow (\forall\epsilon_{n,z}>0) f+\max\sum_{n\in N}\sum_{z\in X_{N\setminus\{n\}}} I_{\{z\}}\epsilon_{n,z} \in \otimes_{n\in N}\mathcal{A}_n \\ &\Rightarrow (\forall\delta>0) f+\delta \in \otimes_{n\in N}\mathcal{A}_n, \end{aligned}$$

whence

$$f \in \mathrm{cl}\,(\otimes_{n\in N}\mathcal{A}_n) \Leftrightarrow (\forall\epsilon_{n,z}>0) f+\sum_{n\in N}\sum_{z\in X_{N\setminus\{n\}}} I_{\{z\}}\epsilon_{n,z} \in \otimes_{n\in N}\mathcal{A}_n.$$

If we use Proposition 6072, we infer that,

$$\begin{aligned} f \in \mathrm{cl}\,(\otimes_{n\in N}\mathcal{A}_n) &\Leftrightarrow (\forall\epsilon_{i,z}>0) f \in \sum_{n\in N}\sum_{z\in X_{N\setminus\{n\}}} I_{\{z\}}\,(\mathcal{A}_n-\epsilon_{i,z}) \\ &\Leftrightarrow f \in \sum_{n\in N}\sum_{z\in X_{N\setminus\{n\}}} I_{\{z\}}\,\mathrm{cl}\,(\mathcal{A}_n) \\ &\Leftrightarrow f \in \otimes_{n\in N}\,\mathrm{cl}\,(\mathcal{A}_n). \end{aligned}$$

□

1.5.2 Symmetry

At the end of this chapter we feel obliged to say a few words about symmetry because this is where the concept of indifference really stands out. We consider a monoid $\mathcal{T}$ of transformations T of $\mathcal{L}$, and some agent's claim that there is symmetry with respect to this monoid. What does that mean? It could be that the agent's dispositions towards the acceptance of gambles stay invariant under transformations of $\mathcal{T}$. This type of symmetry is called **weak symmetry** and is also called **symmetry of the model** by De Cooman and Miranda [15] or **symmetry in beliefs** by Walley [62, Chapter 9]. As the set of indifferent gambles is the workhorse model when it comes to symmetry, we define everything in terms of $(\mathcal{I},\mathcal{D})$ assessments. The definition in terms of sets of acceptable gambles follow at once from Proposition 3_{32}.

Definition 64: *A set of acceptable gambles is* ***weakly symmetric*** *with respect to a monoid of transformations $\mathcal{T}$ if*

$$\mathcal{T}\mathcal{D} \subseteq \mathcal{D} \text{ and } \mathcal{T}\mathcal{I} \subseteq \mathcal{I} \text{ and } \mathcal{T}\mathcal{U} \subseteq \mathcal{U} \text{ and } \mathcal{T}\mathcal{O} \subseteq \mathcal{O},$$

where $\mathcal{T}\mathcal{D} := \{Tf : f \in \mathcal{D} \text{ and } T \in \mathcal{T}\}$, and similarly for the other sets.

This type of symmetry tells something about current views on the modelled variable, but the symmetry will most likely be broken when additional assessments are made, hence the adjective weak. Because the symmetry can break relatively easily we consider this case of less importance—but by no means unimportant—and will not focus on it further.

It could also be possible that the assessor believes that he cannot distinguish between gambles and their transformations. This means that he cannot choose between $f \in \mathcal{L}$ and Tf with T any transformation from the monoid of transformations $\mathcal{T}$. But this means that $f - Tf$ is perceived as an indifferent gamble by the assessor. When additional assessments are made, then either the symmetry will not be broken, or a sure loss situation is created and so making a statement about this type of symmetry is very strong. This type of symmetry is called **strong symmetry** or **model of symmetry** by De Cooman and Miranda [15] or **symmetry in evidence** by Walley [62, Chapter 9].

Definition 65: *A set of acceptable gambles $\mathcal{A}$ is* ***strongly symmetric*** *with respect to a monoid of transformations $\mathcal{T}$ if*

$$\mathcal{I}_{\mathcal{T}} := \operatorname{span}\{f - Tf : f \in \mathcal{L}, T \in \mathcal{T}\} \subseteq \mathcal{I} = \mathcal{A} \cap -\mathcal{A}.$$

▷ **Example 66:** Suppose n different random variables $X_1, X_2, \dots X_n$ representing n different experiments on the same space $\mathcal{X}$. All variables are logically independent and so their joint outcome is defined on the cartesian product space $\mathcal{X}^n$. Given a permutation $\pi: \{1,2,\dots,n\} \to \{1,2,\dots,n\}$, we define the permutation πf of the gamble $f \in \mathcal{L}(\mathcal{X}^n)$ by $\pi f(x_1, x_2, \dots, x_n) \coloneqq f(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)})$. If you claim that a coherent model $\mathcal{A}_{X^n}$ is permutable [62, §9.4] and state that you are predisposed to accept a gamble f, then you are automatically also predisposed to accept any permutation πf of the accepted gamble. If you claim that the model is exchangeable [62, §9.5], [20], then you are willing to change any gamble f for a permuted version πf and you are indifferent towards the gamble $f - \pi f$.

The transformations considered in the literature are usually defined on the underlying possibility space $\mathcal{X}$ rather than on the set of gambles $\mathcal{L}(\mathcal{X})$. This is not a problem as the transformation can in this case be lifted [15] to a linear transformation on the corresponding gamble space as in Example 66. Remark that strong symmetry with respect to a monoid of transformations requires that $\mathcal{I}_{\mathcal{T}} \subseteq \mathcal{I}$ which means that also $\operatorname{span} \mathcal{I}_{\mathcal{T}} \subseteq \mathcal{I}$.

A strong symmetry claim can be considered as a specification of a set of indifferent gambles $\mathcal{I}_{\mathcal{T}}$. If the strong invariance is the only assessment that is made, then the only requirement for this assessment to be valid is that there should be by Equation (1.5) no indifference to sure loss.

Lemma 67 (Amenability of the monoid): *A strong invariance assessment corresponding to a monoid of transformations $\mathcal{T}$ avoids sure loss if and only if*

$$\operatorname{span} \mathcal{I}_{\mathcal{T}} \cap \mathcal{L}_{<0} = \emptyset. \tag{1.41}$$

Equivalently

$$(\forall f_k \in \mathcal{L})(\forall T_k \in \mathcal{T}) \max \sum_k (f_k - T_k f_k) \geq 0. \tag{1.42}$$

Proof: Because by definition, $\mathcal{I}_{\mathcal{T}} \subseteq \mathcal{I}$, we know from ID6$_{30}$ that $\operatorname{span} \mathcal{I}_{\mathcal{T}} \subseteq \mathcal{I}$ and therefore, it should hold by Equation (1.5) that

$$\operatorname{span} \mathcal{I}_{\mathcal{T}} \cap \mathcal{L}_{<0} = \emptyset,$$

which can be rewritten as $(\forall f_k \in \mathcal{L})(\forall T_k \in \mathcal{T})(\sum_k (f_k - T_k f_k) \not< 0)$ which is equivalent to 1.42. □

The term **amenability of a semigroup** was used by Walley following Greenleaf and actually means that there is a $\mathcal{T}$-invariant linear prevision [62, §3.5.6, note 3]. This follows immediately from Propositions 17$_{41}$ and 50$_{65}$. Another interesting problem is the question whether a strong invariance statement is compatible with a given assessment.

Theorem 68 (Dominance theorem): *A strong invariant assessment $\mathcal{I}_{\mathcal{T}}$ is compatible with*[10] *a coherent set of acceptable gambles $\mathcal{A}$ if*

$$\operatorname{span}\mathcal{I}_{\mathcal{T}} \cap (\mathcal{A} \setminus -\mathcal{A}) = \emptyset, \tag{1.43}$$

and

$$(\mathcal{A} + \operatorname{span}\mathcal{I}_{\mathcal{T}}) \cap \mathcal{L}_{<0} = \emptyset. \tag{1.44}$$

Under these conditions, the natural extension $\mathcal{A} + \operatorname{span}\mathcal{I}_{\mathcal{T}}$ of $\mathcal{I}_{\mathcal{T}}$ and $\mathcal{A}$ is coherent and the corresponding lower prevision for a gamble $f \in \mathcal{L}$ can be written as

$$\begin{aligned} \underline{\mathrm{P}}(f) &= \sup\{\underline{E}(f-i) : i \in \operatorname{span}\mathcal{I}_{\mathcal{T}}\}, \\ &= \sup\{\underline{E}_{\mathcal{T}}(f-g) : g \in \mathcal{A}\}, \end{aligned}$$

where

$$\begin{aligned} \underline{E}(h) &:= \sup\{\alpha : h - \alpha \in \mathcal{A}\}, \\ \underline{E}_{\mathcal{T}}(h) &:= \sup\{\alpha : h - \alpha \in \operatorname{ext}(\operatorname{span}\mathcal{I}_{\mathcal{T}})\}, \end{aligned}$$

and $\operatorname{ext}(\operatorname{span}\mathcal{I}_{\mathcal{T}}) = \mathcal{L}_{\geq 0} + \operatorname{span}\mathcal{I}_{\mathcal{T}}$.

Proof: Because a gamble can not be indifferent and desirable at the same time, $\operatorname{span}\mathcal{I}_{\mathcal{T}} \cap \mathcal{A} \setminus -\mathcal{A}$ has to be empty. From ID6$_{30}$ it follows then that $\mathcal{I}_{\mathcal{T}} \cap -\mathcal{A} \setminus \mathcal{A} = \emptyset$ and both assessments $\mathcal{A}$ and $\mathcal{I}_{\mathcal{T}}$ will not be in conflict.

From Equation (1.9) we see—keeping the coherence of $\mathcal{A}$ in mind—that the natural extension of $\operatorname{span}\mathcal{I}_{\mathcal{T}}$ and $\mathcal{A}$ is given by $\operatorname{span}\mathcal{I}_{\mathcal{T}} + \mathcal{A}$ which has to avoid sure loss: $(\operatorname{span}\mathcal{I}_{\mathcal{T}} + \mathcal{A}) \cap \mathcal{L}_{<0} = \emptyset$.

We know that the natural extension of the union of both assessment is given by $\operatorname{span}\mathcal{I}_{\mathcal{T}} + \mathcal{A}$ so the lower prevision of a gamble $f \in \mathcal{L}$ is defined as

$$\begin{aligned} \underline{\mathrm{P}}(f) &= \sup\{\alpha : f - \alpha \in \operatorname{span}\mathcal{I}_{\mathcal{T}} + \mathcal{A}\}, \\ &= \sup\{\alpha : f - i - \alpha \in \mathcal{A} \text{ and } i \in \operatorname{span}\mathcal{I}_{\mathcal{T}}\}, \\ &= \sup\{\sup\{\alpha : f - i - \alpha \in \mathcal{A}\} : i \in \operatorname{span}\mathcal{I}_{\mathcal{T}}\}, \\ &= \sup\{\underline{E}(f-i) : i \in \operatorname{span}\mathcal{I}_{\mathcal{T}}\}. \end{aligned}$$

[10] With "is compatible with" we mean that the assessment $(\mathcal{I}_{\mathcal{T}}, \emptyset, \mathcal{A})$ can be extended to a coherent set of acceptable gambles.

Alternatively, we infer from Axiom A2$_{32}$ and the fact that $0 \in \mathcal{L}_{\geq 0}$, that

$$
\begin{aligned}
\underline{\mathrm{P}}(f) &= \sup\{\alpha\colon f-\alpha \in \operatorname{span}\mathcal{I}_{\mathcal{T}} + \mathcal{A} + \mathcal{L}_{\geq 0}\}, \\
&= \sup\{\alpha\colon f-a-\alpha \in \operatorname{span}\mathcal{I}_{\mathcal{T}} + \mathcal{L}_{\geq 0} \text{ and } a \in \mathcal{A}\}, \\
&= \sup\{\sup\{\alpha\colon f-a-\alpha \in \operatorname{span}\mathcal{I}_{\mathcal{T}} + \mathcal{L}_{\geq 0}\} : a \in \mathcal{A}\}, \\
&= \sup\{\underline{E}_{\mathcal{T}}(f-a)\colon a \in \mathcal{A}\},
\end{aligned}
$$

which concludes the proof. □

Here $\underline{E}_{\mathcal{T}}$ is the smallest and most conservative $\mathcal{T}$-invariant lower prevision, meaning that $\underline{E}_{\mathcal{T}}(f) = 0$ for all $f \in \operatorname{span}\mathcal{I}_{\mathcal{T}}$.

CHAPTER 2

Imprecise probability trees

The legend goes that probability theory sprouted from the correspondence on the “Problème des partis” between Blaise Pascal and Pierre de Fermat. Each of them however, was according to Christiaan Huygens[1] so occupied with “weighty issues” that they kept their findings to themselves, and as a reaction to that, Huygens decided to publish his results on the subject in a manuscript: “Van rekeningh in spelen van geluck”. This manuscript was translated by Van Schooten into Latin and became the first book ever on probability theory: “De ratiociniis in ludo aleae”. Huygens did not claim that the findings were completely his own[2] but one thing that seems to be original is the appearance of what is arguably the first ever published probability tree. The probability tree appeared in an attempt to solve a variation on the Problem of Points (See Figure 2.1).

▷ **Example 69:** The Problem of Points is as follows [26]. Let two players A and B stake equal money on being the first to win n points in a game in which the winner of each point is decided by the toss of a fair coin, heads for A and tails for B. If such a game is interrupted when A still lacks a points and B lacks b, how should the total stakes be divided between them?

[1] In a letter to Franciscus van Schooten jr. (April 27, 1657), Huygens writes: [the French are so occupied with] “swaere questien” [that they] “nochtans elck sijn maniere van uytvinding bedeckt hebben gehouden.”

[2] In the same letter to Van Schooten, Huygens assures that his findings correspond to those of the French.

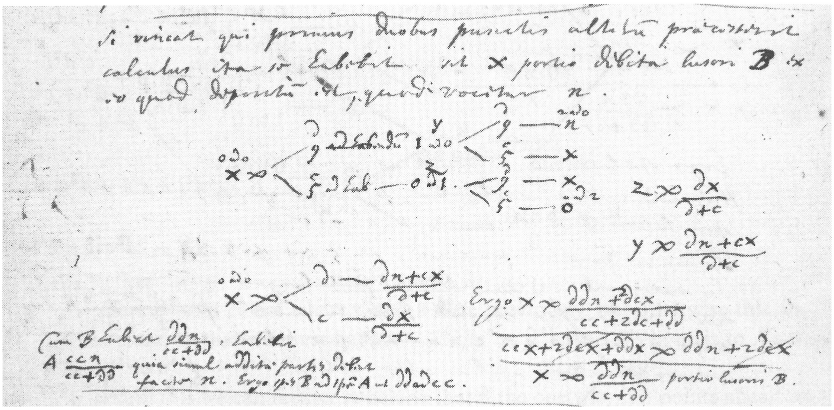


Figure 2.1: Extract from Huygens's manuscript taken from [53] with permission of Glenn Shafer.

What we will do in this chapter is in some sense a revision of Huygens's approach to the Problem of Points, but now, our beliefs about the dice are allowed to be imprecise, and we use conservative reasoning. The question of "what a gamble is worth" remains to us as fundamental as it was in Huygens's solution. Clearly, we intend to develop a framework that can handle a whole variety of problems and not just the Problem of Points. We aim for an account of event-driven random processes.

We start the chapter with the introduction of event trees, which can be seen as probability trees with the probabilities removed. These event trees formalise the possible situations the world can be in. Notation and new concepts will be introduced following Shafer's work on causality [53]. Once the event tree is known, the modelling agent's beliefs about possible transitions between situations have to be incorporated. It is assumed throughout this chapter that the agent only expresses beliefs about situations that follow immediately, using sets of acceptable gambles or lower previsions.

Another theory of uncertainty, where imprecise (lower and upper) probabilities and previsions, rather than precise (or point-valued) probabilities and previsions, have a central part is Glenn Shafer and Vladimir Vovk's game-theoretic account of probability [55]. When comparing Walley's behavioural theory [62] and Shafer and Vovk's game-theoretic framework, they seem to have a rather different interpretation, and they certainly have been influenced by different schools of thought: Walley follows the tradition of Frank Ramsey

[48], Bruno de Finetti [24] and Peter Williams [69] in trying to establish a rational model for a subject's beliefs in terms of her behaviour. Shafer and Vovk follow an approach that has many other influences as well, and is strongly coloured by ideas about gambling systems and martingales. They use Cournot's Principle to interpret lower and upper probabilities (see [54]; and [55, Chapter 2] for a nice historical overview), whereas in Walley's approach, lower and upper probabilities are defined in terms of a subject's betting rates. We have shown in [12] that in many practical situations, the two approaches are strongly connected. This implies that quite a few results, valid in one theory, can automatically be converted and reinterpreted in terms of the other. As an example of this, we will further on prove a generalisation of Hoeffding's inequality.

Although we shall assume below that every local set of acceptable gambles in an imprecise probability tree is only allowed to take values on a finite possibility space, the infinite depth of the tree can result in gambles on an infinite number of situations. As a result, our theory of coherent sets of acceptable gambles, as set out in Chapter 1_{28}, is no longer directly applicable. The main problem is that the maximum of a gamble (or minimum) is not guaranteed to exist. The common way to treat this problem is by redefining the $\prec$ operator: $f \prec g$ if and only if $\sup(f-g) < 0$. The elements in the set $\mathcal{L}_{\prec 0}$ that appear in the avoiding sure loss axiom are the gambles whose supremum is strictly smaller than zero. So when allowing for infinite possibility spaces, $\mathrm{A1}_{32}$ changes slightly. Both $\mathrm{A2}_{32}$ and $\mathrm{A3}_{32}$ remain unchanged. Theorem 7_{35} also generalises in a fairly straightforward fashion. When considering lower (or upper) previsions, there are slight complications:

$$\underline{\mathrm{P}}(f) := \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathcal{A}\},$$

can now also assume infinite values, so $\underline{\mathrm{P}}(f) \in \mathbb{R}^*$ with $\mathbb{R}^* := \mathbb{R} \cup \{-\infty, +\infty\}$. For more details we refer to [21].

2.1 Event trees

Assume that the world, or system, you describe is in a known initial situation and that there is some kind of demigod, which we will call Reality, who repeatedly decides which situation the world is going to be in next. Reality has to follow some plan, however, and the plan is telling him which possible situations it can choose from, in any particular situation. One feature of this plan is that two situations can never evolve into the same situation, which implements the importance of history. This requirement renders Reality's roadmap into a tree where the nodes are the situations and the edges are (Humean) events. This tree is what we call the **event tree**: a roll-out, a roadmap of all possible and relevant situations that Reality can go through. In order to specify the event tree, we see that we need to specify a set of possible situations and some kind of graph that tells us which situations are connected to each other. Also, there is a particular time or causal order, which has to be respected.

2.1.1 Situations and events

In order to define the event tree in a mathematically sound way, we start with the specification of a set $\mathbb{S}$ whose elements we call **situations**. These are all the situations that Reality can possibly be in. On this set $\mathbb{S}$, we assume we have a partial order relation $\sqsubseteq$ that dictates the direction of the evolution through situations. For any two situations $s, t \in \mathbb{S}$ we say that s **precedes** t, or t **follows** s, if $s \sqsubseteq t$ or equivalently, $t \sqsupseteq s$. If $s \sqsubseteq t$ or $t \sqsubseteq s$ then we say that s and t are **ordered**. If s and t are not ordered, we write $s \parallel t$. $s \sqsubset t$ is an abbreviation for $s \sqsubseteq t$ and $s \neq t$ and we call it the "**strictly precedes**" relation; likewise we have the "**strictly follows**" relation $\sqsupset$. The set of all situations that follow a situation t is denoted by $\uparrow t := \{s \in \mathbb{S} : s \sqsupseteq t\}$. Any element of this set—but not t itself—is called a **descendant** or **successor** of t. The set of all situations preceding t is denoted $\downarrow t := \{s \in \mathbb{S} : s \sqsubseteq t\}$. Any situation that strictly precedes t is called an **ancestor** or **predecessor** of t.

We are now able to give the formal definition of an event tree.

Definition 70: *A partially ordered set $\mathbb{S}$ with partial order relation $\sqsubseteq$ is a* ***(discrete) event tree*** $(\mathbb{S}, \sqsubseteq)$ *if it satisfies:*

ET1. *the set $\downarrow t$ of any situation $t \in \mathbb{S}$ is well-ordered by $\sqsubseteq$,*[3]

ET2. *there is some element $\square \in \mathbb{S}$ that precedes all elements of $\mathbb{S}$: $\mathbb{S}$ has a bottom.*

The unique **initial situation** $\square$ that precedes all elements is also called the **root** of the tree. Maximal elements of $\sqsubseteq$ are called the **leaves** of the tree

[3] Remember that a set is well-ordered if it is totally ordered and every nonempty subset has an infimum (See Definition 150_{197} in Appendix B_{194}).

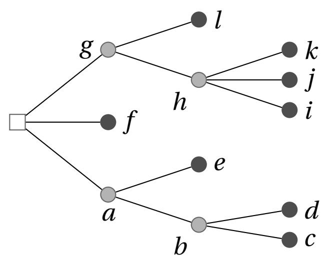


Figure 2.2: A simple event tree for Reality, displaying the initial situation $\square$, other non-terminal situations (such as a, g, h) as grey circles, and terminal situations, (such as f, l, k) as black circles. Observe that $g \sqsubset i$, $g \parallel a$ and $e \sqsupset a$. Observe also $\uparrow a = \{a, b, c, d, e\}$ and $\downarrow a = \{\square, a\}$. This means that the initial situation $\square$ is the only ancestor of a, whereas $\{b, c, d, e\}$ are the descendants of a.

or **terminal situations** as they have no successor. The set of all terminal situations, is denoted by Ω, and is also called the **sample space** of the tree.

From the first requirement ET1, it follows that any two situations can only be unordered if they have no common descendants, which effectively transforms the partial order into a tree. The direction of the order relation $\sqsubseteq$ tells us how Reality can evolve through the situations. Often this evolution can be thought of as an evolution in time. In this sense, ET1 implies (amongst other things) that the branches of the tree diverge as time elapses. In this interpretation, $s \sqsubset t$ means then that the situation s happens before t, which leads to a natural way of formalising causality. This is why Shafer [53] adopted this approach.[4]

Consider the set $\mathrm{desc}(t) := \uparrow t \setminus \{t\}$ of all descendants of t. We prove that this set has minimal elements in non-terminal situations. These minimal elements are called the **children** of t and the set of all children is denoted by $\mathrm{ch}(t)$. We also prove that the children of t are the immediate successors of t. If the chain $\downarrow t \setminus \{t\}$ has a greatest element, it is unique, and called the **mother** of s and denoted by $\mathrm{mo}(s)$. Clearly, if $s \in \mathrm{ch}(t)$, then t is the mother of s, $\mathrm{mo}(s) = t$. For the event tree of Figure 2.2, $\mathrm{ch}(b) = \{c, d\}$ and $\mathrm{mo}(b) = a$.

Proposition 71: *Let t be any situation that is not terminal, i.e.* $\mathrm{desc}(t) \neq \emptyset$. *Then it holds that:*

[4] Total ordering instead of the stronger well-ordering ET1 is required in Shafer's definition of event trees. However, Shafer demands well-ordering when he defines regular event trees.

1. $\mathrm{ch}(t) := \{c \in \mathbb{S} : c$ *is a minimal element of* $\mathrm{desc}(t)\} \neq \varnothing$*,*

2. For every element s of $\mathrm{desc}(t)$*, there is a unique element c of* $\mathrm{ch}(t)$ *such that* $c \sqsubseteq s$*,*

3. No two elements of $\mathrm{ch}(t)$ *are ordered, so* $\mathrm{ch}(t)$ *is an antichain.* [5]

Proof: 1. Consider any (maximal) chain C in $\uparrow t \setminus \{t\}$ and any element s of C. Then $\downarrow s$ is well-ordered by $\mathrm{ET1}_{83}$. As $\downarrow s \cap C$ is a subset of $\downarrow s$, it will be well-ordered as well. This implies that $\downarrow s \cap C$ has a smallest element which is also the smallest element of C. So any chain C in $\mathrm{desc}(t)$ has a smallest element in $\mathrm{desc}(t)$, which implies, by Zorn's lemma, that $\mathrm{desc}(t)$, has minimal elements.

2. Take any $s \in \mathrm{desc}(t)$. Then $\downarrow s \cap \mathrm{desc}(t)$ is well-ordered and therefore has a smallest element c. Clearly, $c \sqsubseteq s$ and we know that c is a minimal element of $\mathrm{desc}(t)$, so belongs to $\mathrm{ch}(t)$.

   Ex absurdo, suppose there is some $c' \in \mathrm{desc}(t)$ such that $c' \sqsubset c$. Then $c' \in \downarrow s \cap \mathrm{desc}(t)$, which contradicts that c is the smallest element of $\downarrow s \cap \mathrm{desc}(t)$.

   Finally, for unicity: suppose $c_1, c_2 \in \mathrm{ch}(t)$ and $c_1 \sqsubseteq s$, $c_2 \sqsubseteq s$. This implies that $c_1, c_2 \in \downarrow s$. As $\downarrow s$ is a chain, $c_1 \sqsubseteq c_2$ or $c_2 \sqsubseteq c_1$ and it follows from 3. that $c_1 = c_2$.

3. Consider any $c_1, c_2 \in \mathrm{ch}(t)$ and assume that $c_1 \sqsubseteq c_2$. But $c_1, c_2 \in \mathrm{desc}(t)$ and are undominated by definition. This implies that $c_1 = c_2$. □

Chains, paths, events and cuts

Shafer [53, §1.7 page 23] makes the distinction between Humean and de Moivrean events where the former correspond to something localised in time and space whereas the de Moivre event corresponds to the usual subset of the sample space. We will adopt the de Moivre interpretation where events are subsets of the sample space. However, some technicalities have to be taken into account when trees with infinite depth are considered. The depth of the tree can be understood as the supremum cardinality of all the chains.[6]

Definition 72: *A* ***path*** $\omega \in 2^{\mathbb{S}}$ *is a chain that is maximal. The set of all paths is denoted by* $\mathbb{P}$*.*

[5] An **antichain** C is a non-empty and unordered subset of $\mathbb{S}$. $(\forall s, t \in C)(s \parallel t)$ (See Definition 149_{195} in Appendix B_{194})

[6] A **chain** C is a non-empty subset of $\mathbb{S}$ that is totally ordered by $\sqsubseteq$. $(\forall s, t \in C)(s \sqsubseteq t$ or $t \sqsubseteq s)$ (See Definition 148_{195} in Appendix B_{194})

A path ω is a chain that is not included in any other chain and the set of all paths that contain or **go through** a situation t is denoted by $\mathbb{P}_t$. Clearly, $\mathbb{P}$ coincides with $\mathbb{P}_\square$. Even if the depth of the tree is infinite, the Axiom of Choice (Hausdorff's Maximal Chain Principle) guarantees that every chain is included in some path.

Every path can be identified with a terminal situation when the tree has a finite depth. When the depth is infinite however, then a path may fail to contain a terminal situation, so paths can no longer be identified with terminal situations. To overcome this problem, a regularity axiom can be added to the definition of the event tree.

Definition 73: *A poset* $(\mathbb{S},\sqsubseteq)$ *is a* ***regular*** *event tree if it is a discrete event tree (satisfies* ET1$_{83}$ *and* ET2$_{83}$*) and if in addition*

ET3. *every chain in* $\mathbb{S}$ *has a greatest element in* $\mathbb{S}$*.*

Remark that this axiom is trivially fulfilled when the depth of the tree is finite. If a tree is not regular then it can always be made regular by adding extra situations to the tree [53]. Therefore, **we assume from now on that every event tree we consider is regular**. In a regular event tree, we can identify paths and terminal situations. We will do so throughout the text.

A (de Moivrean) **event** A is a set of paths $A \subseteq \mathbb{P}$. Because paths can be identified with their terminal situations in a regular event tree, we can also speak without confusion about the event that corresponds to the set of terminal situations $B \subseteq \Omega$ and we will also call this the event B. With an event A, we can associate its **indicator** I_A, which is the real-valued map on Ω that assumes the value 1 on A, and 0 elsewhere.

We denote by $E(t) := \{\omega \in \Omega \colon t \sqsubseteq \omega\}$ the set of all terminal situations that are preceded by t or equivalently, the set of all paths that go through t. $E(t)$ is the event that corresponds to Reality getting to a situation t .

Cuts of a situation

Unless explicitly stated differently, we will assume that the event tree is regular: so, from now on, every path corresponds with a terminal situation.

Definition 74: *Let t be a situation in an event tree* $(\mathbb{S},\sqsubseteq)$*. A* **cut** *U of t is a maximal antichain in* $\uparrow t$*.*

As with paths, Hausdorff's Maximal Chain Principle tells us that we can always extend any set of not-ordered nodes (an antichain) to a cut.

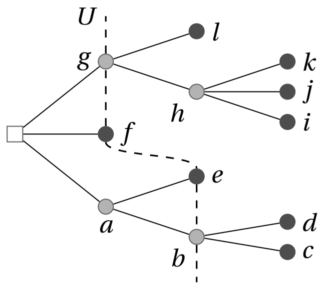


Figure 2.3: A simple event tree for Reality. The set $U := \{g, f, e, b\}$ defines a cut (of $\square$) which precedes the cut $\Omega = \{l, k, j, i, f, e, d, c\}$ of terminal situations: $U \sqsubseteq \Omega$. The set $\{e, b\}$ is not a cut of the tree, but it is a cut of situation a, $\{e, b\} = \mathrm{ch}(a) \in \mathbb{U}_a$.

If a situation $s \sqsupseteq t$ precedes (follows) some element of a cut U of t, then we say that s **precedes** (**follows**) U, and we write $s \sqsubseteq U$ ($s \sqsupseteq U$). Similarly for "strictly precedes (follows)". For two cuts U and V of t, we say that U **precedes** V if each element of U is followed by some element of V. Also, the set Ω of terminal situations is a cut of $\square$, called its **terminal cut**. The event $E(t)$ is the corresponding terminal cut of a situation t.

The next lemma gives some of the more useful properties of cuts and paths.

Lemma 75:

1. *In a regular event tree, a path and a cut intersect in exactly one situation.*

2. *If a set of situations has exactly one situation in common with each cut, then the set is a path.*

3. *If a set of situations has exactly one situation in common with each path, then the set is a cut.*

4. *Every ancestor of an element of a path is again in the same path.*

5. *If s is a situation and U is a cut, then exactly one of the following is true: s is in U, s has a strict descendant in U, or s has a strict ancestor in U.*

The proof can be found in [53, Proposition 11.2].

We can also define the **set of all cuts** of a situation t, which is denoted by $\mathbb{U}_t$. The set $\mathbb{U}_\square$ of all cuts of $\square$ will also be denoted $\mathbb{U}$. The set of all cuts of the situation t that precede the cut U is denoted $\mathbb{U}_t^U$. Shafer [53] proved the following interesting property of the set of cuts of a regular event tree.

Proposition 76: *Consider any situation $t \in \mathbb{S}$ of a regular event tree $(\mathbb{S}, \sqsubseteq)$. Then $(\mathbb{U}_t, \sqsubseteq)$ is a complete lattice.*

As the child of a non-terminal situation t is a situation that immediately follows it, we see that the set $\mathrm{ch}(t)$ of children of t constitutes a cut of t, called its **children cut** (See Proposition 71_{84}). We will also call it the **move space** of Reality in the situation t as it consists of the only situations Reality can evolve to immediately after arriving in t. Although not strictly necessary, we assume that every move space consists of at least two elements, otherwise Reality is confronted with a trivial choice.

2.1.2 Processes and variables

We now have all the necessary tools to represent Reality's possible evolution and have laid the foundations for what can be seen as an **event-driven**, rather than a time-driven, account of a theory of uncertain, or random, processes. The driving events are, of course, the moves that Reality makes. In a theory of processes, we generally consider things that depend on (a succession of) these moves. This leads to the following definitions.

Any (partial) function on the set of situations $\mathbb{S}$ is called a **process**, and any process whose domain includes $\uparrow t$ is called a t-**process**. We will denote processes by capital letters. Of course, a t-process is also an s-process for all $s \sqsupseteq t$ as $\uparrow s \subseteq \uparrow t$; when we call it an s-process, this means that we are restricting our attention to its values in all situations that follow s.

Similarly, any function defined (at least) on the terminal situations following t is called a t-**variable**. A real-valued t-variable can be seen as a gamble on $E(t)$ and will be denoted by a lowercase letter. Given a t-process F, we define the corresponding t-variable $f = \mathrm{var}_{E(t)} F$ on $E(t)$ by:

$$\mathrm{var}_{E(t)} F(\omega) := F(\omega) \quad \text{for any} \quad \omega \in E(t). \tag{2.1}$$

If U is a cut of t, then we call a t-variable f U-**measurable** if for all u in U, f assumes the same value $f(u) := f(\omega)$ for all terminal situations ω

that follow t, $\omega \in E(t)$. In that case we can also consider f as a variable or a gamble on U and will often denote it as f^U or write $f \in \mathcal{L}(U)$. We say that the cut U is a **resolving cut** for f. Given a t-process F and a cut $U \in \mathbb{U}_t$, we define the U-measurable t-variable $\mathrm{var}_U F$ as in Equation (2.1) where now $E(t)$ has been replaced with U.

$$\mathrm{var}_U F(\omega) := F(u) \quad \text{for any } \omega \in E(u) \text{ and } u \in U. \tag{2.2}$$

Consider a t-process F and $U \in \mathbb{U}_t$ any cut of t, then the U-**stopped process** $\mathrm{Stop}_U F$ is the t-process defined by

$$\mathrm{Stop}_U F(s) := \begin{cases} F(s) & s \sqsubseteq U, \\ F(u) & u \sqsubset s \text{ and } u \in U. \end{cases}$$

The cut U is also called a **stopping time**. The corresponding variable is given by $\mathrm{var}_U F$ and is clearly U-measurable:

$$\mathrm{var}_{E(t)} (\mathrm{Stop}_U F) = \mathrm{var}_U F. \tag{2.3}$$

Remark that $\mathrm{var}_{E(t)} F = \mathrm{var}_U F$ does not imply that $F = \mathrm{Stop}_U F$.

Consider a t-process F and $U \in \mathbb{U}_t$ any cut of t, then the U-**killed process** $\mathrm{Kill}_U F$ is the t-process defined by

$$\mathrm{Kill}_U F(s) := \begin{cases} F(s) & s \sqsubseteq U, \\ 0 & u \sqsubset s \text{ and } u \in U. \end{cases}$$

Here

$$\mathrm{var}_U (\mathrm{Kill}_U F) = \mathrm{var}_U F. \tag{2.4}$$

If var, Stop or Kill are applied to a set of processes, we always assume that they are applied pointwise.

▷ **Example 77 (Flipping coins):** Consider flipping two coins, one after the other. This leads to the event tree depicted in Figure 2.4. The identifying labels for the situations should be intuitively clear: e.g., in the initial situation '$\square = ?, ?$' none of the coins have been flipped, in the non-terminal situation '$h, ?$' the first coin has landed 'heads' and the second coin has not yet been flipped, and in the terminal situation 't, t' both coins have been flipped and have landed 'tails'.

First, consider the real process N, which in each situation s, returns the number $N(s)$ of heads obtained so far, e.g., $N(?, ?) = 0$ and $N(h, ?) = 1$. If we restrict the process N to the set Ω of all terminal elements, we get a real variable $n := \mathrm{var}_\Omega N$, whose values are: $n(h, h) = 2$, $n(h, t) = n(t, h) = 1$ and $n(t, t) = 0$.

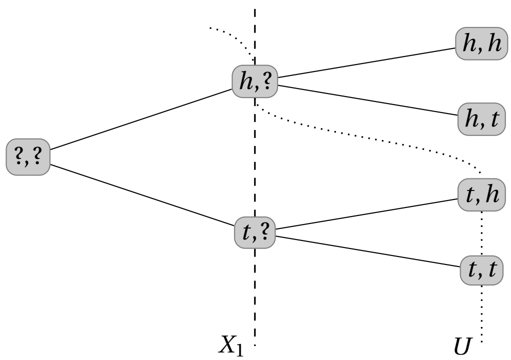


Figure 2.4: The event tree associated with two successive coin flips. Also depicted are two cuts of the initial situation, X^1 and U.

Consider the cut U of the initial situation, which corresponds to the following stopping time: "stop after two flips, or as soon as an outcome is heads"; see Figure 2.4. The values of the corresponding variable n^U are given by: $n^U(h,h) = n^U(h,t) = 1$, $n^U(t,h) = 1$ and $n^U(t,t) = 0$. So n^U is U-measurable, and can therefore be considered as a map on the elements $h,?$ and t,h and t,t of U, with in particular $n^U(h,?) = 1$.

Next, consider the processes $F, G, H \colon \mathbb{S} \to \{h, t, ?\}$, defined as follows:

s	$?,?$	$h,?$	$t,?$	h,h	h,t	t,h	t,t
$F(s)$	$?$	h	t	h	t	h	t
$G(s)$	$?$	h	t	h	h	t	t
$H(s)$	$?$	$?$	$?$	h	t	h	t

F returns the outcome of the latest, G the outcome of the first, and H that of the second coin flip. The associated variables $g := \mathrm{var}_\Omega G$ and $h := \mathrm{var}_\Omega H$ give, in each element of the sample space, the respective outcomes of the first and second coin flips.

The variable g is X_1-measurable: as soon as we reach (any situation on) the cut X_1, its value is completely determined, i.e., we know the outcome of the first coin flip; see Figure 2.4 for the definition of X_1.

We can associate with the process F the variable $f^{X_1} := \mathrm{var}_{X_1} F$ that is also X_1-measurable: it returns, in any element of the sample space, the outcome of the first coin flip. Alternatively, we can stop the process F after one coin flip, which leads to the X_1-stopped process $\mathrm{Stop}_{X_1} F$. This new process is of course equal to G, and for the corresponding variable, we have that $g := \mathrm{var}_\Omega G = \mathrm{var}_\Omega \mathrm{Stop}_{X_1} F = \mathrm{var}_{X_1} F$.

2.2 Imprecise probability trees

Until now we have focussed on the situations Reality can be in and on how Reality can pass through these situations. In this section we introduce a second player, whom we will call Subject and who has beliefs about the actual transition Reality will make once he is in a situation t. We assume that the local belief models are given as coherent sets of acceptable gambles. These local models represent Subject's beliefs when reality is in the initial situation $\square$ about what Reality will do immediately after getting to a particular situation. We call such a model, where the event tree and the immediate prediction beliefs are put together, an imprecise probability tree.

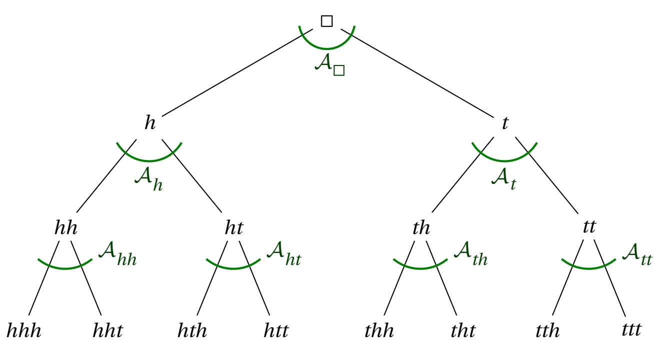
Definition 78: *An* ***imprecise probability tree*** $(\mathbb{S}, \sqsubseteq, \mathcal{A}_\cdot)$ *is an event tree* $(\mathbb{S}, \sqsubseteq)$ *with local model* $\mathcal{A}_s$ *attached to each non-terminal situation* $s \in \mathbb{S} \setminus \Omega$*. A local model* $\mathcal{A}_s$ *in a non-terminal situation* s *is a coherent set of acceptable gambles on the children cut* $\mathrm{ch}(s)$*, conditional on getting to the situation* s*.*

In order to simplify proofs and manoeuvre away the discrepancy between terminal and non-terminal nodes, we assume that a coherent set is assigned to each terminal node as well. More specifically, the coherent set $\mathcal{A}_s$ of a terminal node s consist of all non-negative real numbers $\mathbb{R}_{\geq 0}$:

$$\mathcal{A}_s := \mathbb{R}_{\geq 0} \text{ for all } s \in \Omega. \tag{2.5}$$

A local belief model $\mathcal{A}_t$ **should not be interpreted dynamically**. In other words, $\mathcal{A}_t$ does not stand for the set of acceptable gambles on $\mathrm{ch}(t)$ that Subject accepts after Reality has got to situation t. All beliefs should be fixed and specified explicitly beforehand, in the initial situation $\square$.

▷ **Example 79:** Assume a coin is tossed three times. This means that at every non-terminal situation, either heads h or tails t is concatenated with the current situation.



The event tree is unrolled and in every non-terminal situation s, a local prediction model is given in terms of a coherent sets of acceptable gambles on $\operatorname{ch}(s)$. If the outcome of the first toss was heads for example, then we are in situation h and the beliefs about the possibility of going to one of the next situations ht, hh is thus given by the coherent set of acceptable gambles $\mathcal{A}_h$ defined for all gambles in $\mathcal{L}(\{hh, ht\})$.

From now on, we will consider only a special class of imprecise event trees, which we call **finitary imprecise probability trees**. These are imprecise probability trees that satisfy two extra constraints: 1. every situation has only a finite number of children: $|\operatorname{ch}(s)| \in \mathbb{N}_{\geq 0}$; and 2. there are only a finite number of paths that contain local models with practically impossible events.

2.2.1 Selections and gamble processes

When given an imprecise probability tree, we can consider a special partial process that is not real-valued, but instead returns a gamble for every situation. Such a process is called a **gamble-valued process**. If the returned gamble is acceptable and if, on every path, only a finite number of acceptable gambles are selected that differ from zero, then we call the gamble-valued process a **selection** and denote this process by $\mathcal{S}$. Moreover, when we write $\mathcal{S}(s)(u)$ with $s \sqsubset u$, we mean the acceptable gamble $\mathcal{S}(s)$ selected from $\mathcal{A}_s$ evaluated in the child of s that precedes u (which is unique by Proposition 71_{84}).

Definition 80: *Consider an imprecise probability tree $(\mathbb{S}, \sqsubseteq, \mathcal{A}_\cdot)$ and a situation t. Then we call t-**selection** any t-process $\mathcal{S}$ that satisfies:*

1. $\mathcal{S}(s) \in \mathcal{A}_s$,
2. $\mathcal{S}(s) \neq 0$ *only for a finite number of situations s in each path corresponding to an element of $E(t)$:*

$$(\forall \omega \in E(t)) |\{s \in {\downarrow}\omega : \mathcal{S}(s) \neq 0\}| \in \mathbb{N}_{\geq 0}$$

We call a $\square$-selection simply a selection.

*With a t-selection, we can construct a real-valued t-process $\sum_t \mathcal{S}$, called the **summed** t-**selection process**, whose value in any $s \sqsupseteq t$ is given by*

$$\textstyle\sum_t \mathcal{S}(s) := \sum_{t \sqsubseteq v \sqsubset s} \mathcal{S}(v)(s), \tag{2.6}$$

so $\sum_t \mathcal{S}(t) = 0$.

A selection is allowed to differ from zero in only a finite number of situations in each path. This prevents the sum in Equation (2.6) from diverging. We denote the **set of all t-selections** by $\mathrm{Sels}_t(\mathcal{A}_.)$ and the **set of all summed t-selections** by $\mathrm{SumSels}_t(\mathcal{A}_.)$,

$$\mathrm{SumSels}_t(\mathcal{A}_.) := \left\{\sum_t \mathcal{S} : \mathcal{S} \in \mathrm{Sels}_t(\mathcal{A}_.)\right\}. \tag{2.7}$$

We can also focus on the t-variables $\mathrm{var}_U(\sum_t \mathcal{S})$, $U \in \mathbb{U}_t$, that correspond with the summed t-selections $\sum_t \mathcal{S}$ evaluated in U. For the set $\mathrm{var}_U(\mathrm{SumSels}_t(\mathcal{A}_.))$, we can prove the following useful properties.

Lemma 81:

1. $\mathrm{var}_U(\sum_t \mathcal{S}) = \mathrm{var}_U(\sum_t(\mathrm{Kill}_U \mathcal{S}))$ *for any t-selection* $\mathcal{S} \in \mathrm{Sels}_t(\mathcal{A}_.)$

2. $\mathrm{var}_U(\mathrm{SumSels}_t(\mathcal{A}_.)) = \{\mathrm{var}_U \sum_t \mathcal{S} : \mathcal{S} \in \mathrm{Kill}_U(\mathrm{Sels}_t(\mathcal{A}_.))\}$.

Proof: We know from Equation (2.6) that $\mathrm{Stop}_U \sum_t \mathcal{S} = \sum_t \mathrm{Kill}_U \mathcal{S}$ whence

$$\mathrm{var}_U\left(\sum_t \mathcal{S}\right) = \mathrm{var}_U\left(\mathrm{Stop}_U\left(\sum_t \mathcal{S}\right)\right) = \mathrm{var}_U\left(\sum_t(\mathrm{Kill}_U \mathcal{S})\right)$$

by Equation (2.3).

To prove Property 81.2, it follows from the definition of $\mathrm{SumSels}_t(\mathcal{A}_.)$ and Lemma 81.1 that

$$\begin{aligned}\mathrm{var}_U(\mathrm{SumSels}_t(\mathcal{A}_.)) &= \left\{\mathrm{var}_U\left(\sum_t \mathcal{S}\right) : \mathcal{S} \in \mathrm{Sels}_t(\mathcal{A}_.)\right\}\\ &= \left\{\mathrm{var}_U\left(\sum(\mathrm{Kill}_U \mathcal{S}_t)\right) : \mathcal{S} \in \mathrm{Sels}_t(\mathcal{A}_.)\right\}\\ &= \left\{\mathrm{var}_U\left(\sum_t \mathcal{S}\right) : \mathcal{S} \in \mathrm{Kill}_U(\mathrm{Sels}_t(\mathcal{A}_.))\right\}.\end{aligned}$$ □

Ideally, we would like the set of gambles $\mathrm{var}_{E(t)}(\mathrm{SumSels}_t(\mathcal{A}_.))$, associated with all t-selections, to be coherent. It follows immediately from Equation (2.6) and $\mathrm{A3}_{32}$ that, given $\alpha \in \mathbb{R}_{\geq 0}$,

$$\sum_t(\alpha \mathcal{S}_1 + \mathcal{S}_2) = \alpha \sum_t \mathcal{S}_2 + \sum_t \mathcal{S}_2. \tag{2.8}$$

We infer that $\mathrm{posi}\left(\mathrm{var}_{E(t)}(\mathrm{SumSels}_t(\mathcal{A}_.))\right) = \mathrm{var}_{E(t)}(\mathrm{SumSels}_t(\mathcal{A}_.))$ so $\mathrm{var}_{E(t)}(\mathrm{SumSels}_t(\mathcal{A}_.))$ satisfies Axiom $\mathrm{A3}_{32}$. There is however a problem with Axiom $\mathrm{A2}_{32}$ as $\mathcal{L}_{\geq 0}(E(t)) \notin \mathrm{var}_{E(t)}(\mathrm{SumSels}_t(\mathcal{A}_.))$. The reason for this is that in the summation in Equation (2.6), no selections in terminal situations are incorporated. For this reason, we introduce, for any $U \in \mathbb{U}_t$, the

U-measurable summed t-selections (see also Lemma 83 further on for the justification of the use of "measurable" here):[7]

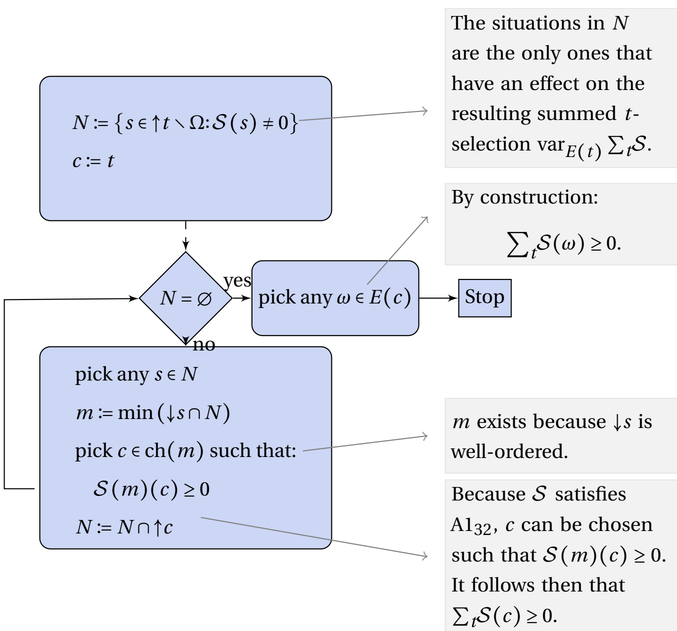
$$\begin{aligned}\mathcal{A}^{U}_{E(t)} &\coloneqq \mathrm{var}_U\left(\mathrm{SumSels}_t(\mathcal{A}_\cdot)\right) + \mathcal{L}_{\geq 0}(U)\\ &= \left\{\mathrm{var}_U \textstyle\sum_t \mathcal{S} \colon \mathcal{S} \in \mathrm{Sels}_t(\mathcal{A}_\cdot)\right\} + \mathcal{L}_{\geq 0}(U).\end{aligned}$$

The set $\mathcal{A}^{E(t)}_{E(t)}$ will also be denoted by $\mathcal{A}_{E(t)}$, and $\mathcal{A}^{\{\omega\}}_{E(\omega)} = \mathcal{A}_\omega = \mathbb{R}_{\geq 0}$, for any terminal situation ω. We now prove that $\mathcal{A}^{U}_{E(t)} \subset \mathcal{L}(U)$ is a coherent set of acceptable gambles on U.

Proposition 82: *Let $(\mathbb{S}, \sqsubseteq, \mathcal{A}_\cdot)$ be an imprecise probability tree, t one of its situations and $U \in \mathbb{U}_t$ a cut of t. Then the set of all U-measurable summed t-selections $\mathcal{A}^{U}_{E(t)}$ is a coherent set of acceptable gambles on $\mathcal{L}(U)$.*

Proof:

A1. As $\mathcal{A}^{U}_{E(t)} \subseteq \mathcal{A}_{E(t)} \coloneqq \mathcal{A}^{E(t)}_{E(t)}$, and $\mathcal{A}_{E(t)} = \mathrm{var}_{E(t)}\left(\mathrm{SumSels}_t(\mathcal{A}_\cdot)\right) + \mathcal{L}_{\geq 0}(E(t))$, it is sufficient to show that $\mathrm{var}_{E(t)}\left(\mathrm{SumSels}_t(\mathcal{A}_\cdot)\right)$ avoids sure loss. We will prove something stronger, namely that, whatever selection $\mathcal{S}$ we take, we can always find a terminal situation ω such that $\sum_t \mathcal{S}(\omega) \geq 0$. So fix a t-selection $\mathcal{S}$ and consider to this end the following algorithm.



[7] Recall from Section 2.1.2$_{88}$ that $\mathcal{L}_{\geq 0}(U)$ with $U \in \mathbb{U}_t$ can be identified with the non-negative gambles on $E(t)$ that are U-measurable.

It is clear that A1$_{32}$ is proved if we can show that the algorithm is guaranteed to stop. Suppose, ex absurdo, that the algorithm does not terminate. This means that there is a sequence $\{c_n, m_n\}_{n\in\mathbb{N}_{\geq 0}}$ such that

$$c_1 \sqsubseteq m_1 \sqsubset c_2 \sqsubseteq m_2 \sqsubset c_3 \sqsubseteq m_3 \sqsubset \dots$$

and $\mathcal{S}(m_k) \neq 0$. But this is clearly in contradiction with the requirement that, on each path, a selection differs from zero in only a finite number of situations (Definition 80$_{92}$).

A2. By definition, $\mathcal{A}^U_{E(t)}$ includes $\mathcal{L}_{\geq 0}(U)$.

A3. If f and g in $\mathcal{A}^U_{E(t)}$ and $\alpha \geq 0$, then we infer from Lemma 81$_{93}$ that there are (U-killed) selections $\mathcal{S}_1$ and $\mathcal{S}_2$, and non-negative gambles $r_1, r_2 \in \mathcal{L}_{\geq 0}(U)$ such that

$$\begin{aligned} \alpha f + g &= \alpha \left(\mathrm{var}_U \textstyle\sum_t \mathcal{S}_1 + r_1\right) + \mathrm{var}_U \textstyle\sum_t \mathcal{S}_2 + r_2 \\ &= \mathrm{var}_U \textstyle\sum_t (\alpha \mathcal{S}_1 + \mathcal{S}_2) + \alpha r_1 + r_2. \end{aligned}$$

where we used Equation (2.8) in the last step. It follows immediately from A3$_{32}$ that $\alpha\mathcal{S}_1 + \mathcal{S}_2$ is again a U-killed selection and obviously, $\alpha r_1 + r_2 \in \mathcal{L}_{\geq 0}(U)$, hence we conclude that $\alpha f + g \in \mathcal{A}^U_{E(t)}$. □

It is not necessarily the case, as pointed out in Section 2.1.2$_{88}$, that a U-measurable variable can be written as the variable resulting from a U-killed selection process (or U-stopped summed selection process). When considering summed t-selections however, we have the following interesting relation, which explains the name we gave earlier to $\mathcal{A}^U_{E(t)}$.

Lemma 83: *$\mathcal{A}^U_{E(t)}$ is the set of all U-measurable summed t-selections.*

$$\mathcal{A}^U_{E(t)} = \left\{ f \in \mathcal{A}_{E(t)} : f \text{ is } U\text{-measurable} \right\}.$$

Proof: Clearly, every $\mathrm{var}_U\left(\sum_t \mathcal{S}\right)$ is U-measurable for any $\mathcal{S} \in \mathrm{Sels}_t(\mathcal{A}.)$.

Conversely, consider any $\mathcal{S} \in \mathrm{Sels}_t(\mathcal{A}.)$ such that $\mathrm{var}_{E(t)}\left(\sum_t \mathcal{S}\right)$ is U-measurable. Then it is sufficient to show that there is some $\beta \in \mathcal{L}_{\geq 0}(U)$ such that

$$\mathrm{var}_{E(t)}\left(\textstyle\sum_t \mathcal{S}\right) = \mathrm{var}_U \left(\textstyle\sum_t \mathrm{Kill}_U \mathcal{S}\right) + \beta.$$

Fix any $u \in U$, then for all $\omega, \omega' \in E(u)$:

$$\mathrm{var}_{E(t)}\left(\textstyle\sum_t \mathcal{S}\right)(\omega) = \sum_{t \sqsubseteq s \sqsubset u} \mathcal{S}(s)(u) + \textstyle\sum_u \mathcal{S}(\omega),$$

$$\mathrm{var}_{E(t)}\left(\textstyle\sum_t \mathcal{S}\right)(\omega') = \sum_{t\sqsubseteq s\sqsubset u} \mathcal{S}(s)(u) + \textstyle\sum_u \mathcal{S}(\omega'),$$

and we infer from the U-measurability of $\mathrm{var}_{E(t)}\left(\sum_t \mathcal{S}\right)$ that $\mathrm{var}_{E(u)}\left(\sum_u \mathcal{S}\right)$ must be a constant map on $E(u)$. We also know that $\mathrm{var}_{E(u)}\left(\sum_u \mathcal{S}\right) \in \mathcal{A}_{E(u)}$ and because $\mathcal{A}_{E(u)}$ avoids sure loss by Proposition 82$_{94}$, we infer that $\mathrm{var}_{E(u)}\left(\sum_u \mathcal{S}\right) = \beta(u) I_{E(u)}$ with $\beta(u) \in \mathbb{R}_{\geq 0}$.

As this reasoning holds for any $u \in U$, we infer that there is a $\beta \in \mathcal{L}_{\geq 0}(U)$ such that

$$\mathrm{var}_{E(t)}\left(\textstyle\sum_t \mathcal{S}\right)(\omega) = \sum_{t\sqsubseteq s\sqsubset u} \mathcal{S}(s)(u) + \beta(u),$$

whence

$$\mathrm{var}_{E(t)}\left(\textstyle\sum_t \mathcal{S}\right) = \mathrm{var}_U\left(\textstyle\sum_t \mathrm{Kill}_U \mathcal{S}\right) + \beta \in \mathcal{A}^U_{E(t)}. \qquad \square$$

Sets of summed t-selections satisfy the following, very interesting property, because it can be seen as a generalised version of marginal extension.

Theorem 84: *Let t be a situation of an imprecise probability tree $(\mathbb{S}, \sqsubseteq, \mathcal{A}_\cdot)$ and consider cuts $U, V \in \mathbb{U}_t$ such that $U \sqsubseteq V$. Then it holds that:*

$$\mathcal{A}^V_{E(t)} = \mathcal{A}^U_{E(t)} + \sum_{u\in U} I_{E(u)} \mathcal{A}^V_{E(u)}.$$

Proof: From Equation (2.6), we infer for any t-selection $\mathcal{S}$ that

$$\begin{aligned}\textstyle\sum_t \mathcal{S} &= \mathrm{Stop}_U \textstyle\sum_t \mathcal{S} + \sum_{u\in U} I_{\uparrow u} \textstyle\sum_u \mathcal{S}|_{\uparrow u} \\ &= \textstyle\sum_t \mathrm{Kill}_U \mathcal{S} + \sum_{u\in U} I_{\uparrow u} \textstyle\sum_u \mathcal{S}|_{\uparrow u}\end{aligned} \tag{2.9}$$

If we define the t-selection $\mathcal{S}^U$ and the u-selections $\mathcal{S}_u$, $u \in U$, by

$$\begin{aligned}\mathcal{S}^U &:= \mathrm{Kill}_U \mathcal{S}, \\ \mathcal{S}_u &:= \mathcal{S}|_{\uparrow u},\end{aligned}$$

then we infer from Equation (2.9) that

$$\begin{aligned}\mathrm{var}_V \textstyle\sum_t \mathcal{S} &= \mathrm{var}_V \textstyle\sum_t \mathcal{S}^U + \sum_{u\in U} I_{E(u)}\, \mathrm{var}_V \textstyle\sum_u \mathcal{S}_u \\ &= \mathrm{var}_U \textstyle\sum_t \mathcal{S}^U + \sum_{u\in U} I_{E(u)}\, \mathrm{var}_V \textstyle\sum_u \mathcal{S}_u,\end{aligned}$$

and it follows from $\mathcal{L}_{\geq 0}(V) = \sum_{u\in U} I_{E(u)} \mathcal{L}_{\geq 0}(\uparrow u \cap V)$ that

$$\mathcal{A}^V_{E(t)} \subseteq \mathcal{A}^U_{E(t)} + \sum_{u\in U} I_{E(u)} \mathcal{A}^V_{E(u)}.$$

To prove that $\mathcal{A}^V_{E(t)} \supseteq \mathcal{A}^U_{E(t)} + \sum_{u\in U} I_{E(u)}\mathcal{A}^V_{E(u)}$ we consider any U-killed t-selection $\mathcal{S}^U$ and any V-killed u-selection $\mathcal{S}_u$, $u \in U$. Then we can define a new t-process $\mathcal{S}$ by

$$\mathcal{S}(s) := \begin{cases} \mathcal{S}^U(s) & \text{if } t \sqsubseteq s \sqsubset U, \\ \mathcal{S}_u(s) & \text{if } u \in U \text{ and } u \sqsubseteq s. \end{cases}$$

$\mathcal{S}$ is again a t-selection since clearly only a finite number of nonzero acceptable gambles is selected on each path through t. Because this result holds for any choice of $\mathcal{S}^U$ and $\mathcal{S}_u$, $u \in U$, and because $\mathcal{L}_{\geq 0}(U) \subseteq \mathcal{L}_{\geq 0}(V)$, it follows at once that $\mathcal{A}^V_{E(t)} \supseteq \mathcal{A}^U_{E(t)} + \sum_{u\in U} I_{E(u)}\mathcal{A}^V_{E(u)}$. □

To conclude, we show that $\mathcal{A}_{E(s)}$ can be interpreted as the set of all summed t-selections, updated on the event $E(s)$ where $s \sqsupseteq t$. When updating was defined (See Definition 24$_{47}$), it was said that updating on a practically impossible event will result in the vacuous model, **unless extra information is given**, which is clearly the case here.

Proposition 85: *The family $\mathcal{A}_{E(t)}$, $t \in \mathbb{S}$ of coherent sets of acceptable gambles satisfies the following properties:*

1. $\mathcal{A}_{E(t)} \rfloor E(s) = \mathcal{A}_{E(s)}$,

2. $\operatorname{marg}_{\operatorname{ch}(s)} \left(\mathcal{A}_{E(t)} \rfloor E(s)\right) = \mathcal{A}_s$,

for all situations $s \sqsupseteq t$ with $-I_{E(s)} \notin \mathcal{A}_{E(t)}$. It can therefore be considered jointly coherent.

Proof: 1. Assume $-I_{E(s)} \notin \mathcal{A}_{E(t)}$ and consider any $f \in \mathcal{L}(E(s))$. Then it follows at once from Lemma 88$_{99}$ further on that

$$I_{E(s)} f \in \mathcal{A}_{E(t)} \Leftrightarrow I_{E(s)} f \in \mathcal{A}_{E(s)}.$$

2. We infer from Lemma 83$_{95}$ that

$$\begin{aligned} \operatorname{marg}_{\operatorname{ch}(s)} \left(\mathcal{A}_{E(t)} \rfloor E(s)\right) &= \operatorname{marg}_{\operatorname{ch}(s)} \left(\mathcal{A}_{E(s)}\right) \\ &= \left\{ f \in \mathcal{A}_{E(s)} : f \text{ is } \operatorname{ch}(s)\text{-measurable} \right\} \\ &= \mathcal{A}^{\operatorname{ch}(s)}_{E(s)} = \mathcal{A}_s. \end{aligned}$$ □

2.2.2 Cut conglomerability

Of course we would like to make inferences on imprecise probability trees by using our extension to Williams's and Walley's theory of **conservative reasoning**. This means that we adopt the rationality criteria for sets of acceptable gambles, cf. Definition 4_{32}. Walley adds an extra "rationality" condition that makes the generalised Bayes rule (Proposition 53_{67}) produce smallest coherent inferences, also on infinite spaces. This condition is called conglomerability. We impose conglomerability with respect to all partitions consisting of exact events.

Definition 86 (Cut conglomerability):

A4. *Given $t \in \mathbb{S}$. Let $U \in \mathbb{U}_t$ and assume that $I_{E(u)} f_u \in \mathcal{A} \subseteq \mathcal{L}(E(t))$, where $f_u \in \mathcal{L}(E(u))$ for all $u \in U$. Then it must hold that $\sum_{u \in U} I_{E(u)} f_u \in \mathcal{A}$.*

Instead of demanding conglomerability for every possible combination of situations, we only ask for conglomerability on cuts, which we believe is the only combination of situations it makes sense to condition on: we agree to condition only on exact events, i.e. events of the form $\uparrow u$ where u is a situation of the tree. If the tree contains only a finite number of situations, then cut conglomerability is a consequence of axiom $\mathrm{A3}_{32}$.

A set of acceptable gambles is coherent if it avoids sure loss, incurs partial gain and when it is closed with respect to the posi operator. In analogy to this posi operator, we can introduce a new operator cccl and rephrase the new axiom, cut conglomerability, as being closed with respect to this operator: given $t \in \mathbb{S}$,

$$\mathrm{cccl}_t \mathcal{A} := \left\{ \sum_{u \in U} I_{E(u)} f_u \colon I_{E(u)} f_u \in \mathcal{A}, f_u \in \mathcal{L}(E(u)) \text{ for all } u \in U \text{ and } U \in \mathbb{U}_t \right\}.$$

So, we say that the set $\mathcal{A} \subseteq \mathcal{L}(E(t))$ is cut conglomerable if it is invariant under cccl:

$$\mathrm{cccl}_t \mathcal{A} \subseteq \mathcal{A}.$$

Considering Theorem 84_{96}, it is not all that surprising that $\mathcal{A}_{E(t)}$ is a cut conglomerable set of acceptable gambles.

Proposition 87: *The coherent set of acceptable gambles $\mathcal{A}_{E(t)}$ is cut-conglomerable.*

Proof: Take any cut $U \in \mathbb{U}_t$ and assume that $I_{E(u)} f_u \in \mathcal{A}_{E(t)}$ for every $u \in U$ where $f_u \in \mathcal{L}(E(u))$. If we fix U, then this means, by Lemma 88, that there are $h_u \in \mathcal{A}_{E(u)}$ and $\alpha_u \in \mathbb{R}_{\geq 0}$ such that $-\alpha_u I_{E(u)} \in \mathcal{A}^U_{E(t)}$ and

$$I_{E(u)} f_u = \begin{cases} I_{E(u)} h_u & \text{if } -I_{E(u)} \notin \mathcal{A}^U_{E(t)}, \\ -\alpha_u I_{E(u)} + I_{E(u)} h_u & \text{if } -I_{E(u)} \in \mathcal{A}^U_{E(t)}. \end{cases} \tag{2.10}$$

Let $S := \left\{u \in U : -I_{E(u)} \in \mathcal{A}_{E(t)}\right\}$, then we infer from Equation (2.10) that

$$\sum_{u \in U} I_{E(u)} f_u = \sum_{s \in S} -\alpha_s I_{E(s)} + \sum_{u \in U} I_{E(u)} h_u \tag{2.11}$$

We know by Proposition 82_{94} that $\mathcal{A}^U_{E(t)}$ is coherent, so it follows from the finitary character of the imprecise probability tree that $|S| \in \mathbb{N}_{\geq 0}$ [see Lemma 89] and therefore $\sum_{s \in S} -\alpha_s I_{E(s)} \in \mathcal{A}^U_{E(t)}$. So we infer from Equation (2.11) and Theorem 84_{96} that indeed

$$\sum_{u \in U} I_{E(u)} f_u \in \mathcal{A}^U_{E(t)} + \sum_{u \in U} I_{E(u)} \mathcal{A}_{E(u)} = \mathcal{A}_{E(t)}. \qquad \square$$

Lemma 88: *Consider $U \in \mathbb{U}_t$ and any $u \in U$ and $f_u \in \mathcal{L}(E(u))$ such that $I_{E(u)} f_u \in \mathcal{A}_{E(t)}$. Then there are $h_u \in \mathcal{A}_{E(u)}$ and $\alpha_u \in \mathbb{R}_{\geq 0}$ such that $-\alpha_u I_{E(u)} \in \mathcal{A}^U_{E(t)}$ and*

$$I_{E(u)} f_u = \begin{cases} I_{E(u)} h_u & \textit{if } -I_{E(u)} \notin \mathcal{A}^U_{E(t)}, \\ -\alpha_u I_{E(u)} + I_{E(u)} h_u & \textit{if } -I_{E(u)} \in \mathcal{A}^U_{E(t)}. \end{cases}$$

Proof: If $I_{E(u)} f_u \in \mathcal{A}_{E(t)}$, then we know by Theorem 84_{96} that there are $h^U \in \mathcal{A}^U_{E(t)}$ and $h_v \in \mathcal{A}_{E(v)}$, $v \in U$ such that

$$I_{E(u)} f_u = I_{E(u)} h^U + \sum_{v \in U} I_{E(v)} h_v.$$

This implies that h_v must be constant and equal to $-h^U(v)$ for all $v \in U \setminus \{u\}$, and since each corresponding $\mathcal{A}_{E(v)}$ avoids sure loss, we find that necessarily $h^U(v) \leq 0$, and that

$$I_{E(u)} f_u = I_{E(u)} [h^U(u) + h_u].$$

So we can assume without loss of generality that $h^U(v) = 0$ for all $v \in U \setminus \{u\}$. [To see this, let $g^U := h^U + \sum_{v \in U \setminus \{u\}} I_{E(v)} h_v$, then as $\sum_{v \in U \setminus \{u\}} I_{E(v)} \in \mathcal{L}_{\geq 0}(U)$ and therefore, by coherence, $g^U \in \mathcal{A}^U_{E(u)}$. Moreover, if we let $g_u := h_u \in \mathcal{A}_{E(u)}$ and $g_v := 0 \in \mathcal{A}_{E(v)}$ for all $v \in U \setminus \{u\}$, then clearly also $I_{E(u)} f_u = I_{E(u)} g^U + \sum_{v \in U} I_{E(v)} g_v$.]

The proof is now complete if we observe that $h^U(u) < 0$ implies, by coherence and the fact that h^U is zero elsewhere, that $-I_{E(u)} \in \mathcal{A}^U_{E(t)}$. So if $-I_{E(u)} \in \mathcal{A}^U_{E(t)}$, then $h^U(u) \geq 0$, and by coherence the result of adding this non-negative constant to h_u will still belong to $\mathcal{A}_{E(u)}$. □

Lemma 89: *Let $(\mathbb{S}, \sqsubseteq, \mathcal{A}_.)$ be a finitary imprecise probability tree, and assume that $-I_{E(u)} \in \mathcal{A}_{E(t)}$ where $u \in U$ and $U \in \mathbb{U}_t$. Then there is a situation $s \sqsubseteq u$ such that $-I_{\{s\}} \in \mathcal{A}_{\text{mo}(s)}$.*

Consequently, in a finitary imprecise probability tree, there can only be a finite number of situations $u \in U$ such that $-I_{E(u)} \in \mathcal{A}_{E(t)}$.

Proof: By A1[32], we may assume without loss of generality that $u \neq t$.

If $-I_{E(u)} \in \mathcal{A}_{E(t)}$, then we know that there is a selection $\mathcal{S} \in \text{Sels}_t(\mathcal{A}_.)$ and a gamble $g \in \mathcal{L}_{\geq 0}(U)$, such that

$$-I_{E(u)} = \text{var}_U \sum\nolimits_t \mathcal{S} + g. \tag{2.12}$$

Choose m to be the largest element of the finite chain

$$\left\{s \in {\downarrow} u \colon \sum\nolimits_t \mathcal{S}(\text{mo}(s)) \geq 0 \text{ and } \sum\nolimits_t \mathcal{S}(s) < 0\right\}.$$

Then we infer from Equation (2.12) that[8]

$$0 = \underbrace{\sum\nolimits_t \mathcal{S}(\text{mo}(m))}_{\geq 0} + \mathcal{S}(\text{mo}(m))(c) + \underbrace{\text{var}_U \sum\nolimits_c \mathcal{S}}_{\in \mathcal{A}^V_{E(c)}} + \underbrace{g|_V}_{\geq 0} \quad \text{for all } c \in \text{sib}(m), \tag{2.13}$$

where we let $V := U \cap {\uparrow} c$. Because $\text{var}_U \sum_c \mathcal{S} \in \mathcal{A}^V_{E(c)}$, and since we know from Proposition 82[94] that $\mathcal{A}^V_{E(c)}$ avoids sure loss, we infer from its proof that there is some situation $v \in V$ such that $\text{var}_U \sum_c \mathcal{S}(v) \geq 0$. If we evaluate Equation (2.13) in this v, then we see that

$$\mathcal{S}(\text{mo}(m))(c) \leq 0 \text{ for all } c \in \text{sib}(m),$$

and by construction

$$\mathcal{S}(\text{mo}(m))(m) = \sum\nolimits_t \mathcal{S}(m) - \sum\nolimits_t \mathcal{S}(\text{mo}(m)) < 0.$$

We know that $\mathcal{S}(\text{mo}(m)) \in \mathcal{A}_{\text{mo}(m)}$ and that $\mathcal{A}_{\text{mo}(m)}$ is coherent, so

$$h := \frac{\mathcal{S}(\text{mo}(m))}{-\mathcal{S}(\text{mo}(m))(m)} \in \mathcal{A}_{\text{mo}(m)}.$$

Since $h \leq -I_{\{m\}}$, we conclude that also $-I_{\{m\}} \in \mathcal{A}_{\text{mo}(m)}$. □

[8] $\text{sib}(m) = \text{ch}(\text{mo}(m)) \setminus \{m\}$ is the set of siblings of m.

2.2.3 Extension to global beliefs

At this moment we are able to solve questions about local gambles, i.e. gambles on the children cut of any situation. The question is how to broaden this to gambles on general cuts. To answer these questions, we have to find a way to combine local beliefs into a global belief model on Ω. Of course we would like to find the smallest joint model that is coherent with the local assessments. In practice this means that the coherence conditions A1$_{32}$-A3$_{32}$ and cut-conglomerability A4$_{98}$ have to be satisfied for the newly formed joint model.

Consider any situation t. What the agent does when specifying the local models that constitute an imprecise probability tree, amounts to an assessment of a set of acceptable gambles $\tilde{\mathcal{A}}_t$ on $E(t)$:

$$\tilde{\mathcal{A}}_t := \bigcup_{s \in \uparrow t} I_{E(s)} \mathcal{A}_s, \tag{2.14}$$

where as usual we have identified gambles on a cut with cut-measurable gambles on the terminal cut. A strategy to find the smallest coherent set lies in applying the posi and cccl operators repeatedly until we find a set that is invariant under both operators. The problem is that posi and cccl do not commute in general:

$$\operatorname{posi} \operatorname{cccl}_t \tilde{\mathcal{A}}_t \neq \operatorname{cccl}_t \operatorname{posi} \tilde{\mathcal{A}}_t,$$

and it is not clear whether this procedure will ever converge, nor whether if it converges, the limits will satisfy A1$_{32}$-A4$_{98}$.

Instead of walking this path, we choose to start from the previously defined set $\mathcal{A}_{E(t)}$ of all selections. We do already know that this set is coherent (A1$_{32}$-A3$_{32}$) by Proposition 82$_{94}$ and cut conglomerable (A4$_{98}$) by Proposition 87$_{98}$ and it clearly contains $\tilde{\mathcal{A}}_t$. It also is jointly coherent with the elements that constitute $\tilde{\mathcal{A}}_t$ by Proposition 85$_{97}$. What is not clear is whether it is the smallest such set. This is what we prove next.

Theorem 90: *Consider a finitary imprecise probability tree $(\mathbb{S}, \sqsubseteq, \mathcal{A}_\cdot)$, and any $t \in \mathbb{S}$. Then the set of acceptable gambles $\mathcal{A}_{E(t)}$ is the smallest set that satisfies*

$$\operatorname{marg}_{\mathrm{ch}(s)} \left(\mathcal{A}_{E(t)} \rfloor E(s)\right) = \mathcal{A}_s \quad \text{for any } s \sqsupseteq t \text{ such that } -I_{E(s)} \notin \mathcal{A}_{E(t)}, \tag{2.15}$$

and that is coherent and cut conglomerable.

Proof: We know from Propositions 82_{94} and 87_{98} that $\mathcal{A}_{E(t)}$ satisfies $\mathrm{A1}_{32}$-$\mathrm{A4}_{98}$. From Proposition 85_{97} we know that $\mathcal{A}_{E(t)}$ is jointly coherent with the local assessments $\mathcal{A}_s$. We only have to prove that $\mathcal{A}_{E(t)}$ is the smallest cut conglomerable, coherent set that is jointly coherent with these local assessments.

Assume, ex absurdo, that there is a coherent and cut conglomerable set of acceptable gambles $\mathcal{A} \subseteq \mathcal{L}(E(t))$ that satisfies Equation (2.15) such that $\mathcal{A}_{E(t)} \nsubseteq \mathcal{A}$. Then we know that there is a gamble $f \in \mathcal{A}_{E(t)} \setminus \mathcal{A}$. This means that there is a t-selection $\mathcal{S}$ and gamble $p \in \mathcal{L}_{\geq 0}(E(t))$ such that

$$f = \mathrm{var}_{E(t)} \sum\nolimits_t \mathcal{S} + p \notin \mathcal{A},$$

so by Lemma 91, we can construct a sequence

$$t =: c_1 \sqsubset c_2 \sqsubset c_3 \sqsubset c_4 \ldots$$

such that $\mathcal{S}(\mathrm{mo}(c_{i+1})) \neq 0$ when $i \in \mathbb{N}_{>0}$. But this means that the t-selection $\mathcal{S}$ differs from zero on an infinite number of situations in the path through $\{c_k\}_{k \in \mathbb{N}_{>0}}$, which contradicts the finitary character of the imprecise probability tree. □

The following lemmas are used when proving Theorem 90.

Lemma 91: *Let $\mathcal{A}$ be a cut conglomerable and coherent set of acceptable gambles on $E(t)$ such that $\mathrm{marg}_{\mathrm{ch}(s)}(\mathcal{A} \rfloor E(s)) = \mathcal{A}_s$ for any $s \sqsupseteq t$ such that $-I_{E(s)} \notin \mathcal{A}$. Consider any t-selection $\mathcal{S}$ and any gamble $p \in \mathcal{L}_{\geq 0}(E(t))$. If there is some situation $c_1 \sqsupseteq t$ such that*

$$\mathrm{var}_{E(t)} \sum\nolimits_{c_1} \mathcal{S} + p \notin \mathcal{A},$$

then there must be some $c_2 \sqsupset c_1$ such that $\mathcal{S}(\mathrm{mo}(c_2)) \neq 0$ and

$$\mathrm{var}_{E(t)} \sum\nolimits_{c_2} \mathcal{S} \notin \mathcal{A}.$$

Proof: Consider the set $S := \{s \sqsupseteq t : \mathcal{S}(s) \neq 0\}$. This set cannot be empty because in that case we would have $\mathrm{var}_{E(t)} \sum_{c_1} \mathcal{S} + p = p \in \mathcal{A}$ by $\mathrm{A2}_{32}$. We then know by Lemma 92 that for any $\omega \in E(t)$:[9]

$$\sum\nolimits_{c_1} \mathcal{S}(\omega) + p(\omega) = p(\omega) + \sum_{m \in \min S} I_{E(m)}(\omega) \left(\mathcal{S}(m)(\omega) + \sum_{c \in \mathrm{ch}(m)} \sum\nolimits_c \mathcal{S}(\omega) \right)$$

[9]Here, $\min S$ is the set of all minimal elements of the partial order $(S, \sqsubseteq)$.

$$
\begin{aligned}
&= p(\omega) + \sum_{m\in\min S} I_{E(m)}(\omega)\mathcal{S}(m)(\omega) \\
&\qquad + \sum_{m\in\min S} I_{E(m)}(\omega) \sum_{c\in\mathrm{ch}(m)} \textstyle\sum_c \mathcal{S}(\omega) \\
&= p(\omega) + \sum_{m\in\min S} I_{E(m)}(\omega)\mathcal{S}(m)(\omega) \\
&\qquad + \sum_{m\in\min S} \sum_{c\in\mathrm{ch}(m)} I_{E(c)}(\omega) \textstyle\sum_c \mathcal{S}(\omega).
\end{aligned}
$$

and therefore

$$
\mathrm{var}_{E(t)} \textstyle\sum_{c_1}\mathcal{S} + p = p + \displaystyle\sum_{m\in\min S} I_{E(m)}\mathcal{S}(m) + \sum_{m\in\min S}\sum_{c\in\mathrm{ch}(m)} I_{E(c)}\,\mathrm{var}_{E(c)} \textstyle\sum_c \mathcal{S}.
$$

Clearly, every gamble $I_{E(m)}\mathcal{S}(m) \in \mathcal{A}$ (see Lemma 93 when $-I_{E(m)} \in \mathcal{A}$), and because $\mathcal{A}$ is cut conglomerable, we infer that $\sum_{m\in\min S} I_{E(m)}\mathcal{S}(m) \in \mathcal{A}$ and by A3$_{32}$ that $p + \sum_{m\in\min S} I_{E(m)}\mathcal{S}(m) \in \mathcal{A}$. But then we know from A3$_{32}$ and $\mathrm{var}_{E(t)} \sum_{c_1}\mathcal{S} + p \notin \mathcal{A}$ that the gamble $\sum_{m\in\min S}\sum_{c\in\mathrm{ch}(m)} I_{E(c)}\,\mathrm{var}_{E(c)}\sum_c\mathcal{S} \notin \mathcal{A}$. Again, because $\mathcal{A}$ is cut conglomerable, this implies that there must be some $m^* \in \min S$ such that $\sum_{c\in\mathrm{ch}(m^*)} I_{E(c)}\,\mathrm{var}_{E(c)}\sum_c\mathcal{S} \notin \mathcal{A}$, and by repeating the cut conglomerability argument for the children cut of this m^*, we know that there must be some $c_2 \in \mathrm{ch}(m^*)$ such that $I_{E(c_2)}\,\mathrm{var}_{E(c_2)}\sum_{c_2}\mathcal{S} \notin \mathcal{A}$. As $\mathrm{mo}(c_2) = m^* \in \min S$, we conclude that $\mathcal{S}(\mathrm{mo}(c_2)) \neq 0$. □

Lemma 92: *Let $S \neq \emptyset$ be a subset of $\mathbb{S}$. Then*

$$
\min S = \bigcup_{s\in S} \min(S \cap \downarrow s)
$$

and $\min S \neq \emptyset$.

Proof: Pick any $s \in S$. Then we know by ET1$_{83}$ that $m := \min(S\cap\downarrow s)$ exists and is unique. Assume ex absurdo that $m \notin \min S$. This means that there is some situation $s^* \in S$ such that $s^* \neq m$ and $s^* \sqsubseteq m$. But then $s^* \in S\cap\downarrow s$ and $s^* \sqsubseteq m$, a contradiction with $m = \min(S\cap\downarrow s)$.

We have proved that $\min S \supseteq \bigcup_{s\in S}\min(S\cap\downarrow s)$ and therefore also that $\min S \neq \emptyset$. To prove that $\min S \subseteq \bigcup_{s\in S}\min(S\cap\downarrow s)$, it is sufficient to notice that $m \in \min S$ implies that $m = \min(S\cap\downarrow m)$. □

Lemma 93: *Let $\mathcal{A} \subseteq \mathcal{L}(E(t))$ be a coherent set of acceptable gambles and assume that $-I_{E(m)} \in \mathcal{A}$. Then $I_{E(m)}\mathcal{L}(E(m)) \subseteq \mathcal{A}$.*

Proof: From A2$_{32}$ we infer that $I_{E(m)}\mathcal{L}_{\geq 0}(E(m)) \subseteq \mathcal{A}$ and from A3$_{32}$ we infer that $-I_{E(m)}\mathbb{R}_{\geq 0} \in \mathcal{A}$. Therefore, it follows from A3$_{32}$ that

$$\mathcal{A} \supseteq I_{E(m)}\mathcal{L}_{\geq 0}(E(m)) - I_{E(m)}\mathbb{R}_{\geq 0} = I_{E(m)}\mathcal{L}(E(m)). \qquad \square$$

2.2.4 Predictive lower and upper previsions

We now use the cut congomerable and coherent set of really desirable gambles $\mathcal{A}_{E(\square)}$ to calculate special lower previsions $\underline{P}(\cdot|t) := \underline{P}(\cdot|E(t))$ in situation $\square$, conditional on an event $E(t)$, i.e., on Reality getting to situation t, as explained in Section 1.4.3$_{60}$.[10] We call such conditional lower previsions **predictive** lower previsions. We then get, using Definition 51$_{66}$, that for any situation t and gamble $f \in \mathcal{L}(\Omega)$,

$$\underline{P}(f|t) := \underline{P}(f|E(t)) = \sup\left\{\alpha \in \mathbb{R} : I_{E(t)}[f-\alpha] \in \mathcal{A}_{E(\square)}\right\} \tag{2.16}$$

$$= \sup\left\{\alpha \in \mathbb{R} : f|_{E(t)} - \alpha \in \mathcal{A}_{E(t)}\right\}, \tag{2.17}$$

where the last transition is a consequence of Theorem 90$_{101}$. We also use the notation $\underline{P}(f) := \underline{P}(f|\square) = \sup\left\{\alpha : f - \alpha \in \mathcal{A}_{E(\square)}\right\}$. It should be stressed that Eq. (2.16) is also valid in terminal situations $t = \omega$, where we let $\underline{P}(f|\omega) = f(\omega)$.

Before we go on, there is an important point that must be stressed and clarified. It is an immediate consequence of Equation (2.17) that when f and g are any two gambles that coincide on $E(t)$, then $\underline{P}(f|t) = \underline{P}(g|t)$. This means that $\underline{P}(f|t)$ is completely determined by the values that f assumes on $E(t)$, and it allows us to define $\underline{P}(\cdot|t)$ on gambles that are only necessarily defined on $E(t)$, i.e., on t-gambles. We will do so freely in what follows.

In the special case of a lower (or upper) prevision of a gamble that is measurable with respect to a cut $U \in \mathbb{U}$, we can simplify Equation (2.16) a bit.

Proposition 94: *Consider an imprecise probability tree $(\mathbb{S}, \sqsubseteq, \mathcal{A}_{\cdot})$ and a cut $U \in \mathbb{U}_t$. Then the lower prevision for the U-measurable t-gamble $f^U \in \mathcal{L}(\Omega)$ are given by*

$$\underline{P}(f^U|t) = \sup\left\{\alpha \in \mathbb{R} : f^U - \alpha \in \mathcal{A}^U_{E(t)}\right\}.$$

[10] We stress again that these are conditional lower previsions on the contingent/updating interpretation, and should not be treated dynamically: they refer to beliefs of a subject at the time Reality is in $\square$, not after Reality has moved to situation t.

Proof: We know that $\underline{P}(f^U|t) = \sup\left\{\alpha \in \mathbb{R}: I_{E(t)}[f^U - \alpha] \in \mathcal{A}_{E(\square)}\right\}$. But clearly,

$$I_{E(t)}[f^U - \alpha] \in \mathcal{A}_{E(t)} \Leftrightarrow f^U - \alpha \in \mathcal{A}_{E(t)} \Leftrightarrow f^U - \alpha \in \mathcal{A}^U_{E(t)}. \qquad \square$$

For any cut U of a situation t, we may define the t-variable $\underline{P}(f|U)$ as the gamble that assumes the value $\underline{P}(f|u)$ in any path ω through a situation $u \in U$. So $\underline{P}(f|U)$ is just the short-hand notation for $\sum_{u\in U} I_{E(u)}\underline{P}(f|u)$. This t-variable is U-measurable by construction, and it can be considered as a variable on U. Observe that this variable may assume values in $\mathbb{R}^*$. In what follows we let $0 \cdot \pm\infty = 0$ by convention.

Proposition 95 (Separate coherence): *Let t be any situation, let U be any cut of t, and let f and g^U be t-gambles, where g^U is U-measurable.*

1. $\underline{P}(E(t)|t) = 1$;
2. $\underline{P}(g^U|U) = g^U$;
3. $\underline{P}(f + g^U|U) = g^U + \underline{P}(f|U)$;
4. *if g^U is moreover non-negative, then* $\underline{P}(g^U f|U) = g^U \underline{P}(f|U)$.

Proof:

1. From Equation (2.16) and from the coherence of $\mathcal{A}_{E(t)}$, we infer that

$$\underline{P}(E(t)|t) = \sup\left\{\alpha \in \mathbb{R}: E(t) - \alpha \in \mathcal{A}_{E(t)}\right\} = \sup\left\{\alpha \in \mathbb{R}: 1 - \alpha \in \mathcal{A}_{E(t)}\right\} = 1.$$

2. This is an immediate consequence of Proposition 94, A3$_{32}$ and Proposition 95.1.
3. $\underline{P}(f + g^U|U) = \sum_{u\in U} I_{E(u)}\underline{P}(f + g^U(u)|u)$ and this is by constant additivity equal to $\sum_{u\in U} I_{E(u)}[\underline{P}(f|u) + g^U(u)] = \underline{P}(f|U) + g^U$.
4. $\underline{P}(g^U f|U) = \sum_{u\in U} I_{E(u)}\underline{P}(g^U(u) f|u)$ and this is by non-negative homogeneity equal to $\sum_{u\in U} I_{E(u)} g^U(u)\underline{P}(f|u) = g^U\underline{P}(f|U)$, because by convention $0 \cdot \pm\infty = 0$. $\square$

2.2.5 Calculating predictive lower prevision using backwards recursion

The Marginal Extension Theorem allows us to calculate the most conservative global belief models $\mathcal{A}_{E(t)}$ that corresponds to the local immediate prediction models $\mathcal{A}_s$. Here beliefs are expressed in terms of sets of acceptable gambles. Can we derive a result that allows us to do something similar for the corresponding lower previsions?

To see what this question entails, first consider a local model $\mathcal{A}_s$: a set of acceptable gambles on $\mathrm{ch}(s)$, where $s \in \mathbb{S} \setminus \Omega$. Using Definition 41_{60}, we can associate with $\mathcal{A}_s$ a coherent lower prevision $\underline{\mathrm{Q}}_s$ on $\mathcal{L}(\mathrm{ch}(s))$. Each gamble g_s on $\mathrm{ch}(s)$ can be seen as an uncertain reward, whose outcome $g_s(c)$ depends on the (unknown) move from situation s to $c \in \mathrm{ch}(s)$ that Reality will make. The **local** (predictive) lower prevision

$$\underline{\mathrm{Q}}_s(g_s) := \sup\{\alpha : g_s - \alpha \in \mathcal{A}_s\} \tag{2.18}$$

for g_s is the supremum acceptable price (for a subject when Reality is in $\Box$) for buying g_s when Reality gets to s.

But as we have seen in Section $2.2.4_{104}$, we can also, in each situation t, derive **global** predictive lower previsions $\underline{\mathrm{P}}(\cdot|t)$ from the global model $\mathcal{A}_{E(t)}$, using Equation (2.16). For each t-gamble f, $\underline{\mathrm{P}}(f|t)$ is **Forecaster**'s (the subject whose beliefs are modelled) inferred supremum acceptable price (in $\Box$) for buying f, contingent on Reality getting to t.

Assume that we are presented with the local predictive lower previsions $\underline{\mathrm{Q}}_s$ instead of sets of acceptable gambles $\mathcal{A}_s$. Is there a way to construct the global predictive lower previsions $\underline{\mathrm{P}}(\cdot|t)$ directly from the local predictive lower previsions $\underline{\mathrm{Q}}_s$? We can infer that there is from the following two theorems, the first of which is merely the lower prevision version of Theorem 84_{96}.

Theorem 96 (Concatenation Formula, Law of Iterated Expectation): *Consider any two cuts U and V of a situation t such that $U \sqsubseteq V$. For all t-gambles f on Ω,*[11]

1. $\underline{\mathrm{P}}(f|t) = \underline{\mathrm{P}}(\underline{\mathrm{P}}(f|U)|t)$;
2. $\underline{\mathrm{P}}(f|U) = \underline{\mathrm{P}}(\underline{\mathrm{P}}(f|V)|U)$.

Proof: It is not difficult to see that the second statement is a consequence of the first, so we only prove the first statement.

a. $\underline{\mathrm{P}}(f|t) \geq \underline{\mathrm{P}}(\underline{\mathrm{P}}(f|U)|t)$. We infer from the super-additivity (Property 48_{64}.2) of the coherent lower prevision $\underline{\mathrm{P}}(\cdot|t)$, that

$$\underline{\mathrm{P}}(f|t) = \underline{\mathrm{P}}(f - \underline{\mathrm{P}}(f|U) + \underline{\mathrm{P}}(f|U)|t)$$

[11] Here, it is implicitly assumed that all expressions are well-defined, e.g., that in the second statement, $\underline{\mathrm{P}}(f|v)$ is a real number for all $v \in V$, making sure that $\underline{\mathrm{P}}(f|V)$ is indeed a gamble.

$$\geq \underline{P}(f - \underline{P}(f|U)|t) + \underline{P}(\underline{P}(f|U)|t),$$

so it suffices to prove that $\underline{P}(f - \underline{P}(f|U)|t) \geq 0$.

Now, for any $\epsilon > 0$, we know that

$$f - \underline{P}(f|U) + \epsilon = \sum_{u \in U} I_{E(u)} \underbrace{[f|_{E(u)} - \underline{P}(f|u) + \epsilon]}_{\in \mathcal{A}_{E(u)}}$$

Hence, it follows from Theorem 84_{96} that $f - \underline{P}(f|U) + \epsilon \in \mathcal{A}_{E(t)}$ for any $\epsilon > 0$. This implies that $\underline{P}(f - \underline{P}(f|U)|t) \geq 0$

b. $\underline{P}(f|t) \leq \underline{P}(\underline{P}(f|U)|t)$. Suppose that $f - \alpha \in \mathcal{A}_{E(t)}$. Then it follows from Theorem 84_{96} that there are $g^U \in \mathcal{A}^U_{E(t)}$ and $g_u \in \mathcal{A}_{E(u)}$, $u \in U$ such that

$$f - \alpha = g^U + \sum_{u \in U} I_{E(u)} g_u.$$

If we apply the lower prevision $\underline{P}(\cdot|u)$ on both sides of the equality we find

$$\underline{P}(f|u) - \alpha = g^U(u) + \underline{P}(g_u|u) \geq g^U(u), \quad u \in U. \tag{2.19}$$

The last inequality is a consequence of g_u being an element of $\mathcal{A}_{E(u)}$. Because Equation (2.19) holds for any $u \in U$, we infer that $\underline{P}(f|U) - \alpha \geq g^U$ so $\underline{P}(f|U) - \alpha \in \mathcal{A}^U_{E(t)}$. Hence $\underline{P}(f|t) \leq \underline{P}(\underline{P}(f|U)|t)$. □

If a t-gamble h is measurable with respect to the children cut $\mathrm{ch}(t)$ of a non-terminal situation t, then we can interpret it as gamble on $\mathrm{ch}(t)$. For such gambles, the following immediate consequence of Proposition 94_{104} tells us that the predictive lower previsions $\underline{P}(h|t)$ are completely determined by the local model $\mathcal{A}_t$.

Proposition 97: *Let t be a non-terminal situation, and consider a* $\mathrm{ch}(t)$*-measurable gamble h. Then $\underline{P}(h|t) = \underline{Q}_t(h)$.*

Proof: Apply Proposition 94_{104} with $U = \mathrm{ch}(t)$. □

These results tells us that all predictive lower (and upper) previsions for imprecise probability trees with finite depth can be calculated using backwards recursion, by starting with the trivial predictive previsions $\overline{P}(f|\Omega) = \underline{P}(f|\Omega) = f$ for the terminal cut Ω, and using only the local models $\underline{P}_t$. This is illustrated in the following simple example.

▷ **Example 98:** Suppose we have $n > 0$ coins. We begin by flipping the first coin: if we get tails, we stop, and otherwise we flip the second coin. Again, we stop if we get tails, and otherwise we flip the third coin, ... In other words, we continue flipping new coins until we get one tails, or until all n coins have been flipped. This leads to the event tree depicted in Figure 2.5. Its sample space is $\Omega = \{t_1, t_2, \dots, t_n, h_n\}$. We

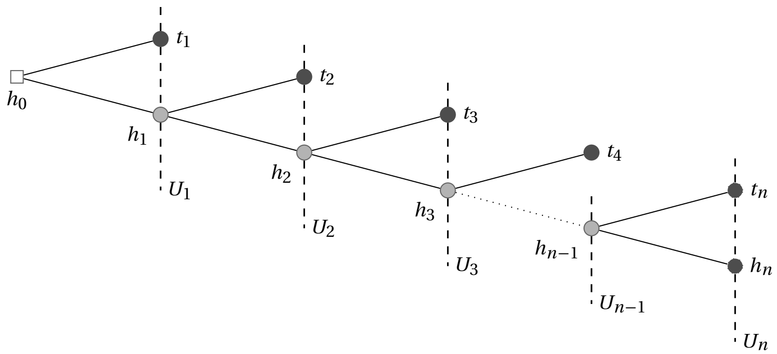


Figure 2.5: The event tree for the uncertain process involving n successive coin flips described in Example 98.

will also consider the cuts $U_1 = \{t_1, h_1\}$ of $\square$, $U_2 = \{t_2, h_2\}$ of h_1, $U_3 = \{t_3, h_3\}$ of h_2, ..., and $U_n = \{t_n, h_n\}$ of h_{n-1}. It will be convenient to also introduce the notation h_0 for the initial situation $\square$.

For the purpose of this example, it will be enough to consider the local predictive lower previsions $\underline{Q}_{h_k}$ on $\mathcal{L}(U_{k+1})$, associated with $\mathcal{A}_{h_k}$ through Eq. (2.18). Forecaster assumes all coins to be approximately fair, in the sense that she assesses that the probability of heads on each flip lies between $\frac{1}{2} - \delta$ and $\frac{1}{2} + \delta$, for some $0 < \delta < \frac{1}{2}$. This assessment leads to the following local predictive lower previsions:[12]

$$\underline{Q}_{h_k}(g) = (1-2\delta)\left[\frac{1}{2}g(h_{k+1}) + \frac{1}{2}g(t_{k+1})\right] + 2\delta\min\{g(h_{k+1}), g(t_{k+1})\}, \quad (2.20)$$

where g is any gamble on U_{k+1}.

Let us see how we can for instance calculate, from the local predictive models $\underline{Q}_{h_k}$, the predictive lower probabilities $\underline{P}(f|s)$ for a gamble f on $E(s)$ and any situation s in the tree. First of all, for the terminal situations it is clear that

$$\underline{P}(\{h_n\}|t_n) = 0 \text{ and } \underline{P}(\{h_n\}|h_n) = 1. \quad (2.21)$$

[12]These so-called linear-vacuous mixtures, or contamination models, are the natural extensions of the probability assessments $\underline{Q}_{h_k}(\{h_{k+1}\}) = \frac{1}{2} - \delta$ and $\overline{Q}_{h_k}(\{h_{k+1}\}) = \frac{1}{2} + \delta$; see Section 1.4.360 and [62, Chapters 3–4] for more details.

We now turn to the calculation of $\underline{P}(\{h_n\}|h_{n-1})$. It follows at once from Proposition 97_{107} that $\underline{P}(\{h_n\}|h_{n-1}) = \underline{Q}_{h_{n-1}}(\{h_n\})$, and therefore, substituting $g = I_{\{h_n\}}$ in Eq. (2.20) for $k = n-1$,

$$\underline{P}(\{h_n\}|h_{n-1}) = \frac{1}{2} - \delta. \tag{2.22}$$

To calculate $\underline{P}(\{h_n\}|h_{n-2})$, consider that, since $h_{n-1} \sqsubseteq U_{n-1}$,

$$\underline{P}(\{h_n\}|h_{n-2}) = \underline{P}(\underline{P}(\{h_n\}|U_{n-1})|h_{n-2}) = \underline{Q}_{h_{n-2}}(\underline{P}(\{h_n\}|U_{n-1}))$$

where the first equality follows from Theorem 96_{106}, and the second from Proposition 97_{107}, taking into account that $g_{n-1} := \underline{P}(\{h_n\}|U_{n-1})$ is a gamble on the children cut U_{n-1} of h_{n-2}. It follows from Eq. (2.21) that $g_{n-1}(t_{n-1}) = \underline{P}(\{h_n\}|t_{n-1}) = 0$ and from Eq. (2.22) that $g_{n-1}(h_{n-1}) = \underline{P}(\{h_n\}|h_{n-1}) = \frac{1}{2} - \delta$. Substituting $g = g_{n-1}$ in Eq. (2.20) for $k = n-2$, we then find that

$$\underline{P}(\{h_n\}|h_{n-2}) = (\frac{1}{2} - \delta)^2. \tag{2.23}$$

Repeating this course of reasoning, we find that more generally

$$\underline{P}(\{h_n\}|h_k) = (\frac{1}{2} - \delta)^{n-k}, \quad k = 0, \dots n-1. \tag{2.24}$$

This illustrates how we can use a backwards recursion procedure to calculate global predictive lower previsions from local ones.

2.3 Some Examples

This section aims at giving a few examples of what can be achieved using imprecise probability trees.

2.3.1 Gambler's ruin

As pointed out in the introduction to this chapter, the solution of the Problem of Points by Huygens was the first published application of a probability tree. Let us consider a slightly modified gambler's ruin problem.

Imagine two players Blaise and Pierre, who repeatedly play rounds of a game that can either be won by Blaise or by Pierre, i.e., there is no tie. As Blaise and Pierre are both slightly addicted to gambling, they decide to play with units of linear utility, called utiles, and the total amount of utiles is equal to $a \in \mathbb{N}_{>0}$. Both players start out with a positive amount of utiles and the player that loses a round has to pay the other player one utile. The game ends when one of the players has no utiles left.

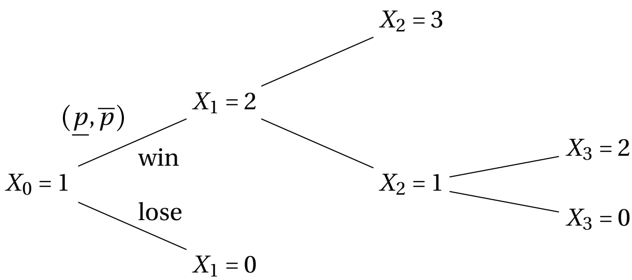


Figure 2.6: Imprecise probability tree representing the gamblers ruin game. Here, Blaise starts with 1 and Pierre with 2 utiles (hence $k = 1$ and $a = 3$) and the game is stopped after three rounds. The probability for Blaise winning in each round lies in the interval $(\underline{p}, \overline{p})$. The random variable X_i stands for the number of utiles of Blaise after i rounds.

If Blaise has at a certain time a capital of k utiles left, what is then a subject's upper probability $\overline{\rho}_k(n)$ of his losing the game after at most n more rounds? It is assumed that the subject believes that the upper probability for Blaise winning a single round is $\overline{p}$ and his lower probability for it is $\underline{p}$, irrespective of the previous outcomes (if Blaise did not win or hit zero before of course).

We know that the upper ruin probability is the conditional upper prevision:

$$\overline{\rho}_k(n) = \overline{\mathrm{P}}(I_{X_n=0} | X_0 = k).$$

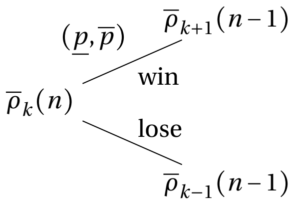
It is clear that the information about the game can be represented by an imprecise probability tree, and the problem is best solved using the Law of Iterated Expectation (see Figure 2.7 and observe that $\overline{\rho}_{k-1}(n-1) \geq \overline{\rho}_{k+1}(n-1)$ by monotonicity):

$$\overline{\rho}_k(n) = \begin{cases} \underline{p}\overline{\rho}_{k+1}(n-1) + (1-\underline{p})\overline{\rho}_{k-1}(n-1) & \text{when } 0 < k < a \text{ and } n > 0, \\ \overline{\rho}_0(n+1) & \text{when } k = 0, \\ \overline{\rho}_a(n+1) & \text{when } k = a, \end{cases}$$

with the initial condition:

$$\overline{\rho}_k(0) = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } k > 0. \end{cases}$$

By letting $n \to \infty$ we get the upper probability $\overline{\rho}_k := \lim_{n\to+\infty} \overline{\rho}_k(n)$ of



If $f \in \mathcal{L}(\{\text{win},\text{lose}\})$ and $f(\text{win}) \leq f(\text{lose})$, then

$$\begin{aligned}\overline{\mathrm{Q}}(f) &= f(\text{win}) + \overbrace{[f(\text{lose}) - f(\text{win})]}^{\geq 0}\overbrace{\overline{\mathrm{Q}}(I_{\{\text{lose}\}})}^{1-\underline{p}}\\ &= \underline{p}f(\text{win}) + (1-\underline{p})f(\text{lose}).\end{aligned}$$

Figure 2.7: Elementary part of the imprecise event tree representing the Gambler's Ruin game.

Blaise losing the game when starting with a capital k by solving the difference equation

$$\overline{\rho}_k = \underline{p}\overline{\rho}_{k+1} + (1-\underline{p})\overline{\rho}_{k-1} \quad ,0 < k < a,$$

with boundary conditions $\overline{\rho}_0 = 1$ and $\overline{\rho}_a = 0$. If we define $\nu := \frac{1-\underline{p}}{\underline{p}}$, then the solution is given by

$$\overline{\rho}_k = \frac{\nu^k - \nu^a}{1-\nu^a}, \quad 0 \leq k \leq a.$$

2.3.2 Forward irrelevance and Wald's equation

A finite collection of independent and identically distributed random variables on a finite space can be considered as one of the most simple stochastic processes there are. However, given the different definitions of independence, immediately there arise problems when trying to generalise this process. In terms of trees, the so-called identical forward irrelevance interpretation is probably the most straightforward generalisation of i.i.d. processes. The generalisation we will consider can be described as an imprecise probability tree where the local models are the same for each non-terminal situation.

Definition 99: *An **identical forward irrelevant process tree** with finite index set $I = \{1,2,\ldots,n\}$ and statespace $\mathcal{X}$, is a finitary imprecise probability tree $(\mathbb{S},\sqsubseteq,\underline{\mathrm{Q}}.)$, where*

1. $\mathbb{S} := \{\square\} \cup \bigcup_{i\in I}\mathcal{X}^i$,

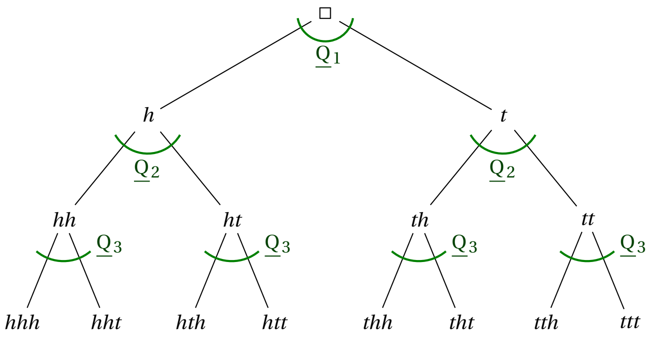


Figure 2.8: A coin is flipped three times. If the lower previsions are such that $\underline{Q}_1(f) = \underline{Q}_2(f) = \underline{Q}_3(f) = \underline{Q}(f)$ for any $f \in \mathcal{L}(\{h,t\})$, then the imprecise probability tree we describe here is a forward irrelevant process tree.

2. $(s \sqsubseteq t) \Leftrightarrow ((s = \square)$ *or* $(s = t)$ *or* $(\exists x \in \mathbb{S} \setminus \{\square\})(sx = t))$,[13]

3. $\underline{Q}_s = \underline{Q}$ *for each* $s \in \mathbb{S} \setminus \Omega$*.*

This set-up can be seen as sampling n times with replacement. If we define the process X by $X(\square) := \square$ and

$$X(sx) := x \quad x \in \mathcal{X} \text{ and } s \in \mathbb{S} \setminus \Omega,$$

and define the cuts $U_i := \{s \in \mathbb{S} : \mathrm{depth}(s) = i\}$,[14] then the variable $X_i := \mathrm{var}_{U_i} X$ (here we make an exception and denote a variable with a capital letter), can be seen as the random variable that corresponds to the outcome of the i-th draw from an urn with possible draws $\mathcal{X}$ (sampling with replacement). It is moreover intuitively clear that

$$X_{1:i-1} \rightarrowtail X_i \quad \text{for all } i \in I$$

whence the sequence $X_1, X_2, \dots, X_n$ shows indeed forward irrelevance (for more details, see [16]).

Proposition 100: *Let* $(\mathbb{S}, \sqsubseteq, \underline{Q}_.)$ *be an identical forward irrelevant process tree with finite index set* $I = \{1, 2, \dots, n\}$ *and statespace* $\mathcal{X}$*. Let* $f \in \mathcal{L}(\mathcal{X})$

[13] Here we mean with sx, the concatenation of s with x. For example if $\mathcal{X} = \{h,t\}$, $s = hht$ and $x = ht$, then $sx = hhtht$.

[14] Here $\mathrm{depth}(s) := |{\downarrow} s| - 1$ stands for the depth of the situation in the tree, the distance from the root node.

and define the process F by $F(\square) := 0$ and $F(sx) := F(s) + f(x)$ for each $s \in \mathbb{S} \setminus \Omega$ and $x \in \mathcal{X}$. Then

$$\underline{P}(\mathrm{var}_\Omega F) = n\underline{Q}(f).$$

Proof: If we apply Theorem 96_{106} with cut $U_{n-1} := \{s \in \mathbb{S} : \mathrm{depth}(s) = n-1\}$, then

$$\begin{aligned}\underline{P}(\mathrm{var}_\Omega F) &= \underline{P}\left(\underline{P}(\mathrm{var}_\Omega F|U_{n-1})\right)\\ &= \underline{P}\left(\underline{P}(\mathrm{var}_{U_{n-1}} F(U_{n-1}) + f|U_{n-1})\right)\\ &= \underline{P}\left(\mathrm{var}_{U_{n-1}} F + \underline{Q}(f)\right) = \underline{P}\left(\mathrm{var}_{U_{n-1}} F\right) + \underline{Q}(f).\end{aligned}$$

Clearly, $\underline{P}(\mathrm{var}_{U_1} F) = \underline{Q}(f)$, and we we have shown by induction that $\underline{P}(\mathrm{var}_\Omega F) = n\underline{Q}(f)$. $\square$

Imagine now that there is another random variable N taking values in $I = \{1, \dots, n\}$, and we have a lower prevision $\underline{Q}_N$ that describes the uncertainty about N. The random sum $\sum_{i=1}^{N} f(X_i)$ can then be reinterpreted as the random stopped process $\mathrm{Kill}_{\tau_N} F$, where

$$\tau_N : I \to \mathbb{U} : \tau_N(n) = \{s \in \mathbb{S} : \mathrm{depth}(s) = n\}.$$

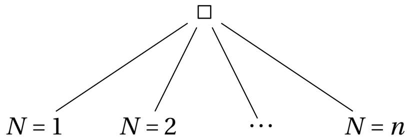
We are now able to formulate a particular equation, known in classical process theory as Wald's equation.

Proposition 101 (Wald's equation): *If N is epistemically irrelevant to the variables $X_1, X_2, \dots, X_n$ linked with the* ***identical*** *forward irrelevant process tree from Proposition 100, then*

$$\underline{P}(\mathrm{var}_{\tau_N} F) = \underline{Q}_N(N\underline{Q}(f)) = \begin{cases}\underline{Q}_N(N)\underline{Q}(f) & \text{if } \underline{Q}(f) \geq 0,\\ \overline{Q}_N(N)\underline{Q}(f) & \text{if } \underline{Q}(f) < 0\end{cases}$$

Proof: The problem can be restated as an imprecise probability tree where the first layer determines the outcome of N, and from the node corresponding with $N = n$ starts the identical forward irrelevant process tree with depth n. So the sets of situation in this new tree are the concatenation of an element of N and the situations of the identical forward irrelevant process tree. Clearly the newly created tree is an imprecise probability tree and it follows immediately from Theorem 84_{96}, Proposition 100 and property $\mathrm{P3}_{65}$ of Proposition 48_{64} that

$$\underline{P}(\mathrm{var}_{\tau_N} F) = \underline{Q}_N\left(\underline{P}(\mathrm{var}_{\tau_N(N)} F|N)\right) = \underline{Q}_N\left(N\underline{Q}(f)\right). \qquad \square$$



2.4 Martingales

The use of martingales—in essence, a summed selection process with an additive offset—is central in Shafer and Vovk's game-theoretic probability [55]. Using their game-theoretic martingales, they are able to derive weak and strong laws of numbers in an elegant manner. The weak laws can be transferred relatively easy to our imprecise probability event tree framework.

Definition 102: *A real-valued t-process $\underline{M}$ is a t-**submartingale** if it can be written as a summed t-selection $\sum_t \mathcal{S}$ plus a constant $\underline{M}(t) \in \mathbb{R}$:*

$$\underline{M} = \underline{M}(t) + \sum_t \mathcal{S}.$$

*A t-process $\overline{M}$ is a t-**supermartingale** if its negation $-\overline{M}$ is a t-submartingale.*

From Equation (2.6) it follows immediately that

$$\mathrm{var}_{\mathrm{ch}(s)} \underline{M} - \underline{M}(s) \in \mathcal{A}_s, \quad \text{for all } s \sqsupseteq t \tag{2.25}$$

if $\underline{M}$ is a t-submartingale. A submartingale is therefore a process that selects an acceptable gamble in each situation. The difference with a summed t-selection process is that there is an initial offset. Whenever an imprecise probability tree has a finite number of situations, Equation (2.25) is also a sufficient condition for the t-process $\underline{M}$ to be a t-submartingale.

In the following lemma, we give an example of a supermartingale that will be useful when proving our version of the weak law of large numbers. The proof given here is a slightly shorter version of the one given in [12, Proof of Theorem 10], which builds on an intriguing idea, used by Shafer and Vovk in a different situation and in a form that is perhaps hard to recognise; see [55, Lemma 3.3].

Lemma 103: *Consider an imprecise probability tree with a finite number of situations.Then, the t-process $\overline{M}$ defined by:*

$$\overline{M}(s) := \overline{M}(t) \prod_{t \sqsubseteq v \sqsubset s} (1 - \xi h_v(s)), \quad \textit{for all } s \sqsupseteq t \tag{2.26}$$

where $\xi \in \mathbb{R}$, $h_s \in \mathcal{A}_s \setminus \{0\}$ and $h_s \leq B$ for all $s \sqsupseteq t$, is a positive t-supermartingale when $\overline{M}(t) > 0$ and $\xi < 1/B$.

Proof: We are dealing with an imprecise probability tree with a finite number of situations, and so we will use Equation (2.25) to prove this lemma. We have to prove for any non-terminal node $s \sqsupseteq t$ that $\mathrm{var}_{\mathrm{ch}(s)} \overline{M} - \overline{M}(s) \in -\mathcal{A}_s$. Let $c \in \mathrm{ch}(s)$, then we infer from the definition of $\overline{M}$ that

$$\begin{aligned}
\overline{M}(c) - \overline{M}(s) &= \overline{M}(t) \prod_{t \sqsubseteq v \sqsubset c} (1 - \xi h_v(c)) - \overline{M}(t) \prod_{t \sqsubseteq v \sqsubset s} (1 - \xi h_v(c)) \\
&= \overline{M}(t) \left(\prod_{t \sqsubseteq v \sqsubset s} (1 - \xi h_v(c)) \right) [1 - \xi h_s(c) - 1] \\
&= -\overline{M}(t) \left(\prod_{t \sqsubseteq v \sqsubset s} (1 - \xi h_v(c)) \right) \xi h_s(c),
\end{aligned}$$

whence

$$\mathrm{var}_{\mathrm{ch}(s)} \overline{M} - \overline{M}(s) = -\underbrace{\overline{M}(t)}_{>0} \underbrace{\xi}_{>0} \left(\underbrace{\prod_{t \sqsubseteq v \sqsubset s} (1 - \xi h_v(c))}_{>0?} \right) \underbrace{h_s}_{\in \mathcal{A}_s} .$$

From A3$_{32}$ we infer that $\overline{M}$ is indeed a supermartingale if $1 - \xi h_v(c) > 0$, for all $t \sqsubseteq v \sqsubseteq s$ and all $s \sqsupseteq t$, so for all $v \sqsupseteq t$. This follows from $\xi < 1/B$ and $1 - \xi h_v \geq 1 - \xi B$. This, together with Equation (2.26), also proves that $\overline{M} > 0$. □

We will also define a particular supermartingale that is going to be useful when proving a generalisation of Hoeffding's inequality. The proof is an adaptation of the proof given by Vovk [58].

Lemma 104: *The t-process $\overline{M}$ defined by:*

$$\overline{M}(s) = \overline{M}(\mathrm{mo}(s))\, e^{-\xi h_{\mathrm{mo}(s)}(s) - \frac{\xi^2 \|h_{\mathrm{mo}(s)}\|_V}{8}}, \quad \textit{for all } s \sqsupset t \tag{2.27}$$

where $\overline{M}(t) = 1$, $\xi \in \mathbb{R}$, $h_s \in \mathcal{A}_s$, $\max h_s \geq 0$ and $\min h_s \leq 0$ for all $s \sqsupseteq t$, is a t-supermartingale.

Proof: Without loss of generality, we can assume that $\|h_s\|_{\mathrm{v}} \neq 0$ as the opposite case represents a trivial supermartingale. To prove the lemma we start from the first -order Taylor expansion of $\alpha(\xi) := \ln(\max h_s e^{-\xi \inf h_s} - \min h_s e^{-\xi \max h_s})$ around zero which is given by

$$\begin{aligned}\alpha(\xi) &= \ln(\overbrace{\max h_s - \min h_s}^{:=\|h_s\|_{\mathrm{v}}}) + 0\xi \\ &\quad + \int_0^{\xi} x \frac{\partial^2}{\partial x^2} \ln(\max h_s e^{-x \min h_s} - \min h_s e^{-x \max h_s})\,\mathrm{d}x.\end{aligned}$$

Working out the second derivative, we get

$$\|h_s\|_{\mathrm{v}}^2 \frac{(\max h_s e^{-x \min h_s})(-\min h_s e^{-x \max h_s})}{(\max h_s e^{-x \min h_s} - \min h_s e^{-x \max h_s})^2},$$

so the integrand of the Lagrange remainder becomes of the form $x\|h_s\|_{\mathrm{v}}^2 p(1-p)$ with $p \in [0,1]$. Therefore, the second derivative is smaller or equal than $x\|h_s\|_{\mathrm{v}}^2/4$. Substituting this upper bound in the integrand yields the inequality

$$\ln(\max h_s e^{-\xi \min h_s} - \min h_s e^{-\xi \max h_s}) \leq \ln(\|h_s\|_{\mathrm{v}}) + \frac{\xi^2 \|h_s\|_{\mathrm{v}}^2}{8},$$

and therefore also

$$\ln\Big(\frac{\max h_s e^{-\xi \min h_s} - \min h_s e^{-\xi \max h_s}}{\|h_s\|_{\mathrm{v}}}\Big) \leq \frac{\xi^2 \|h_s\|_{\mathrm{v}}^2}{8}.$$

By taking the exponential of both sides, we get

$$\frac{\max h_s}{\|h_s\|_{\mathrm{v}}} e^{-\xi \min h_s} + \frac{-\min h_s}{\|h_s\|_{\mathrm{v}}} e^{-\xi \max h_s} \leq e^{\frac{\xi^2 \|h_s\|_{\mathrm{v}}^2}{8}}.$$

and adding $\frac{h_s}{\|h_s\|_{\mathrm{v}}} e^{-\xi \max h_s} - \frac{h_s}{\|h_s\|_{\mathrm{v}}} e^{-\xi \min h_s}$ on both sides, we find:

$$\begin{aligned}&\overbrace{\frac{\max h_s - h_s}{\|h_s\|_{\mathrm{v}}}}^{(1-\eta)} e^{-\xi \min h_s} + \overbrace{\frac{h_s - \min h_s}{\|h_s\|_{\mathrm{v}}}}^{\eta} e^{-\xi \max h_s} \\ &\qquad \leq e^{\frac{\xi^2}{8}\|h_s\|_{\mathrm{v}}} + \frac{h_s}{\|h_s\|_{\mathrm{v}}} \underbrace{(e^{-\xi \max h_s} - e^{-\xi \min h_s})}_{e^{-\xi \min h_s}(e^{-\xi\|h_s\|_{\mathrm{v}}}-1)}.\end{aligned}$$

If we define $\eta := \frac{h_s - \min h_s}{\|h_s\|_{\mathrm{v}}}$, then we see that $\eta \in [0,1]$, and the left-hand side of the inequality above is a convex mixture. As the exponential is a convex function, and because $(1-\eta)\min h_s + \eta \max h_s = h_s$, we get

$$e^{-\xi h_s} \leq e^{\frac{\xi^2 \|h_s\|_{\mathrm{v}}^2}{8}} + \frac{h_s}{\|h_s\|_{\mathrm{v}}} e^{-\xi \min h_s}(e^{-\xi \|h_s\|_{\mathrm{v}}} - 1),$$

which can be rewritten as

$$c_s h_s := \underbrace{\frac{\overbrace{e^{-\xi \min h_s - \frac{\xi^2 \|h_s\|_{\mathrm{v}}^2}{8}}}^{>0} (1 - \overbrace{e^{-\xi \|h_s\|_{\mathrm{v}}}}^{\in(0,1)})}{\underbrace{\|h_s\|_{\mathrm{v}}}_{>0}}}_{>0} h_s \leq 1 - e^{-\xi h_s - \frac{\xi^2 \|h_s\|_{\mathrm{v}}^2}{8}}.$$

As $c_s > 0$ the gamble $c_s h_s$ belongs, by $\mathrm{A3}_{32}$, to the set of acceptable gambles $\mathcal{A}_s$. Consequently, if we multiply both sides with $\overline{M}(s)$ we get

$$\overline{M}(s) - \mathrm{var}_{\mathrm{ch}(s)} \overline{M} \in \mathcal{A}_s.$$

It then follows from Equation (2.25) that $-\overline{M}$ is a submartingale, and consequently $\overline{M}$ is a supermartingale. □

Given the strong relation between selections and submartingales, it cannot come as a surprise that we can express the predictive lower prevision in terms of submartingales, and upper previsions in terms of supermartingales.

Proposition 105: *For all situations t, all $U \in \mathbb{U}_t$ and any U-measurable t-gamble f^U, it holds that*

$$\underline{\mathrm{P}}(f^U|t) = \sup\left\{\underline{M}(t) : \underline{M} \text{ is a } t\text{-submartingale and } f^U \geq \mathrm{var}_U \underline{M}\right\} \quad (2.28)$$

$$\overline{\mathrm{P}}(f^U|t) = \inf\left\{\overline{M}(t) : \overline{M} \text{ is a } t\text{-supermartingale and } f^U \leq \mathrm{var}_U \overline{M}\right\} \quad (2.29)$$

Proof: From Proposition 94_{104} we know that the lower prevision $\underline{\mathrm{P}}(f^U|t)$ of the U-measurable t-variable f^U is given by

$$\begin{aligned} \underline{\mathrm{P}}(f^U|t) &= \sup\left\{\alpha : f^U - \alpha = \mathrm{var}_U \textstyle\sum_t \mathcal{S} + p \text{ where } \mathcal{S} \in \mathrm{Sels}_t(\mathcal{A}.) \text{ and } p \in \mathcal{L}_{\geq 0}(U)\right\}, \\ &= \sup\left\{\alpha : f^U \geq \alpha + \mathrm{var}_U \textstyle\sum_t \mathcal{S} \text{ where } \mathcal{S} \in \mathrm{Sels}_t(\mathcal{A}.)\right\}, \\ &= \sup\left\{\underline{M}(t) : f^U \geq \mathrm{var}_U \underline{M} \text{ where } \underline{M} \text{ is a } t\text{-submartingale}\right\}, \end{aligned}$$

where the last step follows from Definition 102_{114} with $\alpha = \underline{M}(t)$.

Equation (2.29) follows immediately from $\underline{\mathrm{P}}(f) = -\overline{\mathrm{P}}(-f)$. □

2.4.1 Imprecise concentration inequalities

In what follows we call **Forecaster** the agent, who, in □, has certain beliefs about the moves that Reality will make. In this section we will prove a generalised version of the weak law of large numbers and of the Hoeffding - Azuma inequality. Classically, the first is proved using the Chebyshev inequality which in its turn is based on the Markov inequality. These two inequalities generalise fairly easily to imprecise probabilities. Assume that $\epsilon > 0$, then it follows from the monotonicity of $\underline{\mathrm{P}}$ that

$$\epsilon I_{|h|\geq\epsilon} \leq |h| \Rightarrow \overline{\mathrm{P}}(|h| \geq \epsilon) \leq \frac{\overline{\mathrm{P}}(|h|)}{\epsilon},$$

which is the generalisation of the Markov inequality. This inequality implies in its turn the generalised Chebyshev inequality:

$$\overline{\mathrm{P}}(|h| \geq \epsilon) \leq \frac{\overline{\mathrm{P}}(h^2)}{\epsilon^2}.$$

In going from the Chebyshev inequality to the (weak) Law of Large Numbers, the gamble h is typically assumed to be $h = \frac{1}{n}\sum_{i=1}^{n}(f_i - \underline{\mathrm{P}}(f_i))$, where $f_i \in \mathcal{L}(X_i)$ with $\mathcal{X}_i =: \mathcal{X}$, and $f_i(x) =: f(x)$. It is moreover assumed that all X_i are independent and identically distributed. In the imprecise probabilistic framework, assuming that $\underline{\mathrm{P}}_{X_i}(f_i) = \underline{\mathrm{P}}(f)$ and denoting $\frac{1}{n}\sum_{i=1}^{n} f_i$ by $\overline{f}_n$, this results in

$$\overline{\mathrm{P}}\left(\left|\overline{f}_n - \underline{\mathrm{P}}(f)\right| \geq \epsilon\right) \leq \frac{\overline{\mathrm{P}}\left(\left(\overline{f}_n - \underline{\mathrm{P}}(f)\right)^2\right)}{\epsilon^2}.$$

It is now left to prove that $\overline{\mathrm{P}}\left(\left(\overline{f}_n - \underline{\mathrm{P}}(f)\right)^2\right) \to 0$ as $n \to \infty$. This should be done—assuming epistemic independence—using the independent natural extension, which is by no means a trivial affair.

Instead of following the i.i.d. path, we choose to follow the martingale approach. Consider therefore an imprecise probability tree with **finite depth**, with non-terminal situation t and a cut U of t. Define the t-variable n^U such that $n^U(\omega)$ is the distance $d(t,u) := |\{s \in \mathbb{S} : t \sqsubseteq s \sqsubset u\}|$, measured in moves along the tree, from t to the unique situation u in U that $\downarrow\omega$ goes through. n^U is clearly U-measurable, and $n^U(u)$ is simply the distance $d(t,u)$ from t to u. We assume that $n^U(u) > 0$ for all $u \in U$, or in other words that $U \neq \{t\}$. Of course, n^U is bounded because the tree has finite depth, and we denote its minimum by n.

Now consider for each s between t and U a bounded gamble h_s and a real number m_s such that $h_s - m_s \in \mathcal{A}_s$, meaning that Forecaster in situation $\square$ accepts to buy h_s for m_s, contingent on Reality getting to situation s. Let $B > 0$ be any common upper bound for $\max h_s - \min h_s$, for all $t \sqsubseteq s \sqsubset U$. It follows from the coherence of $\mathcal{A}_s$ that $m_s \leq \max h_s$. To make things interesting, we also assume that $\min h_s \leq m_s$, because otherwise $h_s - m_s \geq 0$ and accepting this gamble represents no real commitment on Forecaster's part. As a result, we see that $|h_s - m_s| \leq \max h_s - \min h_s \leq B$.

We are interested in the following t-gamble G^U, given by

$$G^U = \frac{1}{n^U} \sum_{t \sqsubseteq s \sqsubset U} I_{E(s)}[h_s - m_s],$$

which provides a measure for how much, on average, the gambles h_s yield an outcome above Forecaster's accepted buying prices m_s, along segments of the tree starting in t and ending right before U. In other words, G_U measures the average gain for Forecaster along segments from t to U, associated with commitments Forecaster has made and is taken up on, because Reality has to move along these segments.

We would like to study Forecaster's beliefs (in the initial situation $\square$ and contingent on Reality getting to t) in the occurrence of the event

$$\Delta_\epsilon := \left\{G^U \leq -\epsilon\right\} := \left\{\omega \in E(t) : G^U(\omega) \leq -\epsilon\right\},$$

where $\epsilon > 0$. In other words, we want to know $\overline{\mathrm{P}}(\left\{G^U \leq -\epsilon\right\}|t)$, which is Forecaster's infimum rate for selling the bet on the event that his average gain from t to U will be at most $-\epsilon$, contingent on Reality's getting to t. We will generalise two concentration inequalities that are known in classical literature, to wit, the weak Law of Large Numbers and the Hoeffding - Azuma inequality.

Theorem 106 (Weak law of large numbers): *For all* $0 < \epsilon < B$ *and any cut* $U \sqsupset t$*,*

$$\overline{\mathrm{P}}(\left\{G^U \leq -\epsilon\right\}|t) \leq \exp\left(-\frac{n\epsilon^2}{4B^2}\right),$$

where $n := \min_{u \in U} n^U(u)$ *and* $B > 0$ *any common upper bound for* $|h_s - m_s|$.

Proof: From the definition of upper previsions in terms of martingales (Proposition 105_{117}), we know that the proof is done if we can find a t-supermartingale $\overline{M}$ such that $\overline{M}(u) \geq I_{\Delta_\epsilon}(u)$ for all $u \in U$ and $\overline{M}(t) \leq e^{-\frac{n\epsilon^2}{4B^2}}$. In Lemma 103_{114} we defined a positive supermartingale $\overline{M}$ that can be rescaled with $\overline{M}(t)$.

Because $\overline{M} > 0$, we only have to concentrate on these values u in U for which $G^U(u) \leq -\epsilon$. For these u, we have that

$$\begin{aligned}\overline{M}(u) \geq 1 &\Leftrightarrow \overline{M}(t) \prod_{t\sqsubseteq v\sqsubset u} (1-\xi(h_v(u)-m_v)) \geq 1\\ &\Leftrightarrow \sum_{t\sqsubseteq v\sqsubset u} \ln(1-\xi(h_v(u)-m_v)) \geq -\ln\overline{M}(t).\end{aligned}$$

If we assume that $\xi < \frac{1}{2B}$ (remember that ξ should be strictly smaller than $\frac{1}{B}$), then $-\xi(h_v(u)-m_v) > -\xi B > -1/2$. Since $\ln(1+x) \geq x - x^2$ for $x > -\frac{1}{2}$, we then infer for any $u \in U$ such that $G^U(u) \leq -\epsilon$

$$\begin{aligned}\overline{M}(u) \geq 1 &\Leftarrow \sum_{t\sqsubseteq v\sqsubset u} -\xi(h_v(u)-m_v) - \xi^2(h_v(u)-m_v)^2 \geq -\ln\overline{M}(t)\\ &\Leftrightarrow -\xi n^U(u)G^U(u) - \xi^2 \sum_{t\sqsubseteq v\sqsubset u}(h_v(u)-m_v)^2 \geq -\ln\overline{M}(t)\\ &\Leftarrow \xi n^U(u)\epsilon - \xi^2 n^U(u)B^2 \geq -\ln\overline{M}(t)\\ &\Leftrightarrow e^{-\xi n^U(u)\left(\epsilon-\xi B^2\right)} \leq \overline{M}(t)\\ &\Leftarrow e^{-\xi n\left(\epsilon-\xi B^2\right)} \leq \overline{M}(t).\end{aligned}$$

In the last transition we assumed that $\epsilon - \xi B^2 > 0$.

We thus see that

$$\overline{\mathrm{P}}(\left\{G^U \leq -\epsilon\right\}|t) \leq \min_{0<\xi<\frac{1}{2B}} e^{-\xi n\left(\epsilon-\xi B^2\right)} = e^{-\frac{n\epsilon^2}{4B^2}}.$$

The minimum is attained for $\xi = \frac{\epsilon}{2B^2} < \frac{1}{2B}$ and in this case $\epsilon - \xi B^2 = \frac{\epsilon}{2} > 0$ as assumed. □

Theorem 107 (Hoeffding's inequality): *For all $\epsilon > 0$ and any cut $U \sqsupset t$,*

$$\overline{\mathrm{P}}(\left\{G^U \leq -\epsilon\right\}|t) \leq \exp\left(-\frac{2n^2\epsilon^2}{V_U}\right),$$

where $V_U := \max_{u\in U} v^U(u)$ with $v^U(u) := \sum_{t\sqsubseteq s\sqsubset u} \|h_s\|_v^2$.

Proof: From the expression of upper previsions in terms of supermartingales (Equation 2.29$_{117}$), we infer that the theorem is proved if we can find a supermartingale $\overline{M}$, such that $\overline{M}(t) \leq \exp\left(-\frac{2n\epsilon^2}{V_U}\right)$ and such that $\mathrm{var}_U\,\overline{M} \geq I_{G^U \geq -\epsilon}$, or in other words

$$(\forall u \in U)\begin{cases}\overline{M}(u) \geq 1 & \text{if } G^U(u) \leq -\epsilon,\\ \overline{M}(u) \geq 0 & \text{otherwise.}\end{cases}$$

From Lemma 104[115] we know one supermartingale $\overline{M}$ which starts at $\overline{M}(t) = 1$ and obtains

$$\overline{M}(u) = \exp\left(-\xi n^U(u)\, G^U(u) - \frac{\xi^2 v^U(u)}{8}\right),$$

for all $u \in U$. Consequently—as $\overline{M} \geq 0$—we only have to examine the values of this supermartingale on the event Δ_ϵ. On this event we have $G^U \leq -\epsilon$ and therefore for every $\xi > 0$:

$$\overline{M}(u) \geq \exp\left(\xi n^U(u)\epsilon - \frac{\xi^2 v^U(u)}{8}\right),$$

hence, for every $u \in U$ such that $G^U \leq -\epsilon$,

$$\overline{M}(u) \geq \exp\left(\xi\epsilon \underbrace{\min n_U}_{=n} - \frac{\xi^2 \overbrace{\max v^U}^{=V^U}}{8}\right).$$

So,

$$\exp\left(-\xi\epsilon n + \frac{\xi^2 V^U}{8}\right)\overline{M}(u) \geq 1.$$

Now, $\exp\left(-\xi\epsilon n + \frac{\xi^2 V^U}{8}\right)\overline{M}(u)$ is again a supermartingale, with value $e^{-\xi\epsilon n + \frac{\xi^2 V^U}{8}}$ in t and therefore,

$$\overline{\mathrm{P}}(\left\{G^U \leq -\epsilon\right\}|t) \leq \min_{\xi>0} \exp\left(-\xi\epsilon n + \frac{\xi^2 V_U}{8}\right) = \exp\left(-\frac{2n^2\epsilon^2}{V_U}\right). \qquad \square$$

CHAPTER 3

Imprecise Markov chains

One convenient way to model uncertain dynamical systems mathematically is to describe them as Markov chains. Such systems have been studied in great detail, and their properties are very well known. However, in many practical situations, it remains a challenge to accurately identify the transition probabilities in the chain, mainly because the information we may obtain about physical systems is often imprecise and uncertain. As a result, describing a real-life dynamical system as a Markov chain will often lead to unwarranted precision, and the model may therefore jump to conclusions that are not supported by the available information.

For this reason, it seems quite useful to perform probabilistic robustness studies, or sensitivity analyses, for Markov chains. This is especially relevant in decision-making applications. It should come as no surprise, then, that many researchers [27, 31, 43, 65] in Markov Chain Decision Making—inspired by Satia & Lave [49]'s original work [49]—have paid attention to this issue of "imprecision" in Markov chains.

Early work on the more mathematical aspects of modelling such "imprecision" in Markov chains was done by Hartfiel [28] and Kozine & Utkin [36]. Armed with linear programming techniques, Kozine & Utkin [36] also performed an experimental study of the limit behaviour of Markov chains with uncertain transition probabilities. More recently, Škulj [59] has begun a formal study of the time evolution and limit behaviour of such systems.

All these approaches have at least one thing in common: they use **sets of probabilities** to deal with the imprecision in the transition probabilities.

When these probabilities are not well known, they are assumed to belong to certain sets, and robustness analyses are performed by allowing the transition probabilities to vary over such sets. However, this approach leads to a number of computational as well as technical mathematical difficulties. These difficulties can be overcome by tackling the same problem from another angle, one that we know already. Rather than working with sets of transition probabilities, or equivalently, sets of Markov chains, we can consider an imprecise Markov chain, which is a special type of imprecise probability tree.

A large part of this Chapter studies the limit behaviour of stationary imprecise Markov chains, or alternatively, studies the limit behaviour of the **upper transition operator** (see Section 3.2$_{129}$) that corresponds to an imprecise Markov chain. This upper transition operator inherits the properties of upper previsions and consequently, it is a non-expansive map. A very useful result for non-expansive maps by Sine [56, Theo. 1] and Nussbaum [45, 46][1] states that for every element f of the finite-dimensional domain of a non-expansive transformation T, there is some natural number p such that the sequence $\mathrm{T}^{np}f$ converges. More importantly, Sine proves that we can find a finite 'period' p **common** to all maps f on the domain $\mathcal{X}$. This means that, for any f, the set $\omega_{\overline{\mathrm{T}}}(f)$ of limit points of the set of iterates $\left\{\overline{\mathrm{T}}^n f \colon n \in \mathbb{N}_{\geq 0}\right\}$ has a number of elements $\left|\omega_{\overline{\mathrm{T}}}(f)\right|$ that divides this p. $\overline{\mathrm{T}}$ is cyclic on $\omega_{\overline{\mathrm{T}}}(f)$, with period $\left|\omega_{\overline{\mathrm{T}}}(f)\right|$ (and therefore also with period p). Lemmens and Scheutzow [37, Theo. 5.2] managed to prove that an upper bound for the common periods of all topical functions (i.e. monotone and constant additive functions) $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^n$ is $\binom{n}{\lfloor n/2 \rfloor}$. This upper bound is tight in the sense that there is always at least one topical function that has this bound as its smallest common period. However, Akian and Gaubert [1, Cor. 5.6] have shown that for convex maps that are monotone and non-expansive, this bound is equal to the maximal order of the permutation group. This is given by Landau's function g for which $\ln g(n) \sim c_1 \sqrt{n \ln n}$, whereas $\ln \binom{n}{n/2} \sim c_2 n$, for some constants $c_1, c_2 > 0$, as $n \to \infty$.

In Section 3.6$_{143}$ we use these ideas to introduce ergodicity for upper transition operators, and to explain its link with so-called Perron-Frobenius conditions. That there is such a link has already been established by Akian and Gaubert [1, Theo. 1.1] for a more general class of operators. The goal of Akian

[1] Nussbaum found and closed a gap in Sine's argument.

and Gaubert was to determine combinatorial bounds for the orbit lengths of order preserving, convex and sup-norm non-expansive maps. These upper bounds involve the notion of a critical graph. It is shown in [1, Theo. 6.6], that these bounds are tight when the map is piecewise affine. Moreover, in this case, Akian and Gaubert give an algorithm to compute the critical graph. In this paper, we assume in addition to the general assumptions made by Akian and Gaubert, that the map is non-negatively homogeneous and we address the case where all periodic orbits have length one. For this class of maps, we show that the piecewise affine condition can be dispensed with for the critical graph bound to be tight. In Section 3.10$_{154}$ we compare our approach to the critical graph method used by Akian and Gaubert.

3.0.1 A short analysis of classical Markov chains

Consider a finite Markov chain in discrete time, where at consecutive times $n = 1, 2, 3, \dots, N$, $N \in \mathbb{N}_{>0}$ the **state** X_n of a system can assume any value in a finite set $\mathcal{X}$. Here, N is the time horizon. The time evolution of such a system can be modelled as if it traverses a path in an **event tree**. An example of such a tree for $\mathcal{X} = \{a, b\}$ and $N = 3$ is given in Figure 3.1.

The **situations**, or nodes, of the tree have the form $x_{1:k} := (x_1, \dots, x_k) \in \mathcal{X}^k$, $k = 0, 1, \dots, N$. For $k = 0$ there is some abuse of notation as we let $\mathcal{X}^0 := \{\square\}$, where $\square$ is the so-called **initial situation**, or root of the tree. In the cuts $\mathcal{X}^n$ of $\square$, the value of the state X_n at time n is revealed.

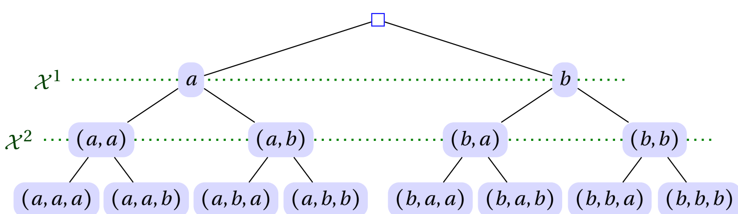


Figure 3.1: The event tree for the time evolution of system that can be in two states, a and b, and can change state at time instants $n = 1, 2$. Also depicted are the respective cuts $\mathcal{X}^1$ and $\mathcal{X}^2$ of $\square$ where the states at times 1 and 2 are revealed.

In a classical analysis, it is generally assumed that we have: (i) a probability distribution over the initial state X_1, in the form of a probability mass function m_1 on $\mathcal{X}$; and (ii) for each situation $x_{1:n}$ that the system can be in at time n, a probability distribution over the next state X_{n+1}, in the form of a probability mass function $q(\cdot|x_{1:n})$ on $\mathcal{X}$. This means that the event tree's immediate prediction model in each non-terminal situation $x_{1:n}$ of the event tree is a **local** probability model telling us about the probabilities of each of its child nodes. This turns the event tree into a so-called **probability tree**; see Shafer [53, Chapter 3] and Kemeny & Snell [35, § 1.9].

The probability tree for a Markov chain is special, because the **Markov Condition** states that when the system jumps from state $X_n = x_n$ to a new state X_{n+1}, the belief model for where the system goes to will only depend on the state $X_n = x_n$ the system was in at time n, and not on its states $X_k = x_k$ at previous times $k = 1, 2, \dots, n-1$. In other words:

$$q(\cdot|x_{1:n}) = q_{n+1}(\cdot|x_n), \quad x_{1:n} \in \mathcal{X}^n,\ n = 1, \dots, N-1, \tag{3.1}$$

where $q_{n+1}(\cdot|x_n)$ is some probability mass function on $\mathcal{X}$. The Markov chain may be non-stationary, as the transition probabilities on the right-hand side in Equation (3.1) are allowed to depend explicitly on the time n. Figure 3.2 gives an example of a probability tree for a Markov chain with $\mathcal{X} = \{a, b\}$ and $N = 3$.

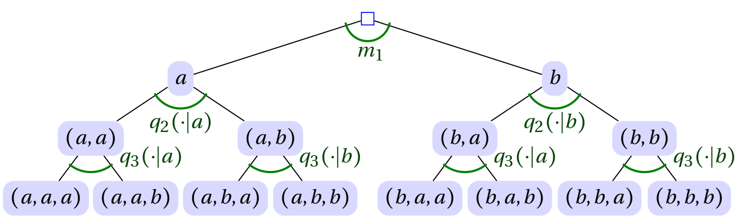


Figure 3.2: The probability tree for the time evolution of a Markov chain that can be in two states, a and b, and can change state at each time instant $n = 1, 2$.

With the local probability mass functions m_1 and $q_{n+1}(\cdot|x_n)$ we associate the linear real-valued **prevision functionals** Q_1 and $\mathrm{Q}_{n+1}(\cdot|x_n)$, given, for all real-valued maps h on $\mathcal{X}$, by

$$\mathrm{Q}_1(h) := \sum_{x_1 \in \mathcal{X}} h(x_1) m_1(x_1) \tag{3.2}$$

$$\mathrm{Q}_{n+1}(h|x_n) := \sum_{x_{n+1} \in \mathcal{X}} h(x_{n+1}) q_{n+1}(x_{n+1}|x_n) \tag{3.3}$$

Throughout, we will formulate our results using previsions, rather than probabilities.

In any probability tree, probabilities and expectations can be calculated very efficiently using backwards recursion. Suppose that in situation $x_{1:n}$, we want to calculate the conditional expectation $\mathrm{P}(f|x_{1:n})$ of some real-valued map f on $\mathcal{X}^N$ that may depend on the values of the states $X_1, \dots, X_N$. Let us indicate briefly how this is done, also taking into account the simplifications due to the Markov Condition (3.1).

To express these simplifications, a prominent part will be played by the so-called **transition operators**[2] T_n and $\mathbb{T}_n$. Consider the linear space $\mathcal{L}(\mathcal{X})$ of all real-valued maps on $\mathcal{X}$. Then the linear operator (transformation) $\mathrm{T}_n: \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ is defined by

$$\mathrm{T}_n h(x_n) := \mathrm{Q}_{n+1}(h|x_n) = \sum_{x_{n+1} \in \mathcal{X}} h(x_{n+1}) q_{n+1}(x_{n+1}|x_n) \tag{3.4}$$

for all real-valued maps h on $\mathcal{X}$. In other words, $\mathrm{T}_n h$ is the real-valued map on $\mathcal{X}$ whose value $\mathrm{T}_n h(x_n)$ in $x_n \in \mathcal{X}$ is the conditional expectation of the random variable $h(X_{n+1})$, given that the system is in state x_n at time n. More generally, we also consider the linear maps $\mathbb{T}_n$ from $\mathcal{L}(\mathcal{X}^{n+1})$ to $\mathcal{L}(\mathcal{X}^n)$, defined by

$$\begin{aligned} \mathbb{T}_n f(x_{1:n}) &:= \mathrm{T}_n(f(x_{1:n}, \cdot))(x_n) \\ &= \mathrm{Q}_{n+1}(f(x_{1:n}, \cdot)|x_n) = \sum_{x_{n+1} \in \mathcal{X}} f(x_{1:n}, x_{n+1}) q_{n+1}(x_{n+1}|x_n) \end{aligned} \tag{3.5}$$

for all $x_{1:n} \in \mathcal{X}^n$ and all real-valued maps f on $\mathcal{X}^{n+1}$.

We begin our illustration of backwards recursion by calculating $\mathrm{P}(f|x_{1:n})$ for the case $n = N-1$. Here

$$\begin{aligned} \mathrm{P}(f|x_{1:N-1}) &= \mathrm{P}(f(x_{1:N-1}, \cdot)|x_{1:N-1}) \\ &= \sum_{x_N \in \mathcal{X}} f(x_{1:N-1}, x_N) q(x_N|x_{1:N-1}) \\ &= \sum_{x_N \in \mathcal{X}} f(x_{1:N-1}, x_N) q_N(x_N|x_{N-1}) = \mathbb{T}_{N-1} f(x_{1:N-1}), \end{aligned} \tag{3.6}$$

where the third inequality follows from the Markov Condition (3.1), and the fourth from Equation (3.5). Using similar arguments for $n = N-2$, we derive

[2] The operators T_n are also called the **generators** of the Markov process; see Whittle [66].

from the Law of Iterated Expectations[3] that

$$\mathrm{P}(f|x_{1:N-2}) = \mathrm{P}(\mathrm{P}(f(x_{1:N-2},\cdot,\cdot)|x_{1:N-2},\cdot)|x_{1:N-2}) = \mathbb{T}_{N-2}\mathbb{T}_{N-1}f(x_{1:N-2}). \tag{3.7}$$

Repeating this argument leads to the backwards recursion formulae

$$\mathrm{P}(f|x_{1:n}) = \mathbb{T}_n\mathbb{T}_{n+1}\ldots\mathbb{T}_{N-1}f(x_{1:n}) \tag{3.8}$$

for $n = 1,\ldots,N-1$, while for $n = 0$, we get

$$\mathrm{P}(f) := \mathrm{P}(f|\square) = \mathrm{Q}_1(\mathbb{T}_1\mathbb{T}_2\ldots\mathbb{T}_{N-1}f). \tag{3.9}$$

In these formulae, f is any real-valued map on $\mathcal{X}^N$.

For instance, if we let f be the indicator functions $I_{\{x_{1:N}\}}$ of the singletons $\{x_{1:N}\}$, Formulae (3.8) and (3.9) allow us to calculate the joint probability mass function p defined by $p(x_{1:N}) = \mathrm{P}(I_{\{x_{1:N}\}})$ for all the variables $X_1,\ldots,X_N$. We can also use them to find the conditional mass functions $p_{n+1}(\cdot|x_n)$ and $p(\cdot|x_{1:n})$ defined by $p_{n+1}(x_{n+1:N}|x_n) = p(x_{n+1:N}|x_{1:n}) = \mathrm{P}(I_{\{x_{1:N}\}}|x_{1:n})$.

3.1 Towards imprecise Markov chains

The treatment above shows that a classical Markov chain can be seen as a special type of event tree with precise probability models attached to the non-terminal nodes. From the previous chapter we already know how to implement this type of model, where local prediction models are used that are more general than linear previsions. This special type of imprecise probability tree is exactly what we define to be an imprecise Markov chain. In this chapter we will make one concession to the generality of this type of Markov chain in that we will assume that the models are given as coherent upper previsions (and not as coherent sets of acceptable gambles). The reason we focus on upper- rather than lower previsions will become clear in Section 3.5135, but the main reason is that there is always a positive upper transition probability to go to a next state, which is not true for lower probabilities.

Definition 108: *An* ***imprecise Markov chain*** *of length N with state space $\mathcal{X}$ and local conditional models $\overline{\mathrm{Q}}_{i+1}(\cdot|X_i)$ on $\mathcal{L}(X_{i+1})$ with $i \in \{1,\ldots,N-1\}$ and initial model $\overline{\mathrm{Q}}_1$ on $\mathcal{L}(X_1)$ is an imprecise probability tree with set of situations $\mathbb{S} = \mathcal{X}_{1:N}$ where*

[3] Also known as the Rule of Total Expectation, or the Rule of Total Probability, or the Conglomerative Property; see, e.g., Whittle [66, § 5.3] or de Finetti [24].

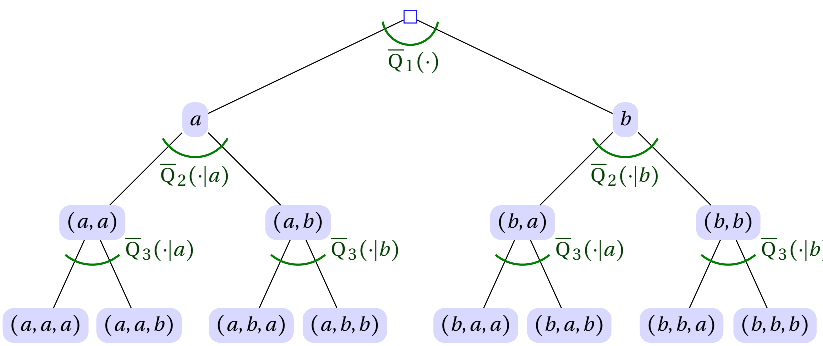
1. *in every non-terminal situation there is a choice amongst one of the elements of* $\mathcal{X}$

$$\mathrm{ch}(s) = \mathcal{X} \quad \textit{for all } s \in \mathbb{S} \setminus \Omega,$$

2. *the local conditional models satisfy the* ***Markov Condition***

$$\overline{\mathrm{Q}}_{i+1}(\cdot|x_{1:i}) = \overline{\mathrm{Q}}_{i+1}(\cdot|x_i) \quad \textit{for all } x_{1:i} \in \mathcal{X}^i \textit{ and } i = 1,2,\ldots,N-1. \tag{3.10}$$

The second condition demands that every variable X_{i+1} should be epistemically irrelevant to its non-parents non-descendants given its parent X_i. This type of irrelevance structure will also provide the interpretational basis for the more general Markov trees discussed in the next chapter. The Markov Condition also justifies the name "state space" as the current state summarises everything there is to know about possible trajectories of the system under study. A classical, or **precise**, Markov chain is an imprecise one with local upper previsions that are linear.



▷ **Example 109:** An imprecise Markov chain of length three, interpreted as an imprecise probability tree. At each step, a transition of an element of the state space $\mathcal{X} = \{a,b\}$ to the same state space $\mathcal{X}$ is possible. An important property of a Markov chain is that the local prediction model depends only on the last state. For example, the local prediction models in (b,b) and in (a,b) are both equal to $\overline{\mathrm{Q}}_3(\cdot|b)$. If the imprecise Markov chain is stationary, then $\overline{\mathrm{Q}}_2(\cdot|b) = \overline{\mathrm{Q}}_3(\cdot|b)$.

If the local conditional models do not depend on the actual index: $\overline{\mathrm{Q}}_{i+1}(\cdot|\cdot) = \overline{\mathrm{Q}}(\cdot|\cdot)$, then we say that the imprecise Markov chain is **stationary** and we can summarise it by the tuple $(\mathcal{X}, \overline{\mathrm{Q}}_1, \overline{\mathrm{Q}}, N)$.

3.2 Upper transition operators

Theorem 96_{106} allows us to comfortably calculate upper previsions of arbitrary gambles on $\mathcal{X}^n$. To streamline notation and provide more insight, we introduce upper transition operators in analogy with transition operators in the precise case.

Definition 110: *The **upper transition operator** $\overline{\mathrm{T}}_i$ of an imprecise Markov chain with local conditional upper previsions $\overline{\mathrm{Q}}_{i+1}$ is given by*

$$\overline{\mathrm{T}}_i\colon \mathcal{L}(X_{i+1}) \to \mathcal{L}(X_i)\colon \overline{\mathrm{T}}_i f(x) = \overline{\mathrm{Q}}_{i+1}(f|X_i = x)$$

So the transition operator $\overline{\mathrm{T}}_i$ is a map, a transformation from $\mathcal{L}(\mathcal{X})$ to $\mathcal{L}(\mathcal{X})$: it takes a gamble $f \in \mathcal{L}(X_{i+1})$ and turns it into a gamble on X_i. By definition, $\overline{\mathrm{T}}_i I_{\{y\}}(x)$ is equal to $\overline{\mathrm{Q}}_{i+1}(I_{\{y\}}|x) = \overline{\mathrm{Q}}_{i+1}(\{y\}|x)$ which is the upper probability to go from the state x at "time point" i to state y at the next time point. What is important to notice, is that for two different states $x, y \in \mathcal{X}$, there is no interaction between $\overline{\mathrm{T}}_i f(x)$ and $\overline{\mathrm{T}}_i f(y)$, i.e. if for example $\overline{\mathrm{T}}_i f(x)$ takes a certain value, then this tells you nothing about $\overline{\mathrm{T}}_i f(y)$. We say that $\overline{\mathrm{T}}_i$ is **rectangular**.[4] More generally, we also consider the maps $\overline{\mathbb{T}}_i$ from $\mathcal{L}(\mathcal{X}^{i+1})$ to $\mathcal{L}(\mathcal{X}^i)$, defined by

$$\overline{\mathbb{T}}_n f(x_{1:n}) := \big(\overline{\mathrm{T}}_n f(x_{1:n},\cdot)\big)(x_n) = \overline{\mathrm{Q}}_{n+1}(f(x_{1:n},\cdot)|x_n) \tag{3.11}$$

for all $x_{1:n}$ in $\mathcal{X}^n$ and all real-valued maps f on $\mathcal{X}^{n+1}$. Of course, we can also consider lower expectations and lower transition operators, which are related to the upper expectations and upper transition operators by conjugacy.

As is the case for coherent upper previsions, it is possible to introduce the notion of an upper transition operator directly, by basing it on a number of defining properties, rather than by referring to an underlying imprecise Markov chain.

Definition 111 (Upper transition operator): *Consider a non-empty finite set of states $\mathcal{X}$, and the set $\mathcal{L}(\mathcal{X})$ of all real-valued maps on $\mathcal{X}$. A transformation $\overline{\mathrm{T}}$ of $\mathcal{L}(\mathcal{X})$ is called an **upper transition operator** if and only if it satisfies the following properties:*

[4]This is inspired by the following analogy: a set is said to be rectangular if it is the Cartesian product of subsets of a set $\mathcal{S}$.

$\overline{\mathrm{T}}$1. $I_{\mathcal{X}} \min h \leq \overline{\mathrm{T}}h \leq I_{\mathcal{X}} \max h$ [boundedness]

$\overline{\mathrm{T}}$2. $\overline{\mathrm{T}}(h_1 + h_2) \leq \overline{\mathrm{T}}h_1 + \overline{\mathrm{T}}h_2$ [sub-additivity]

$\overline{\mathrm{T}}$3. $\overline{\mathrm{T}}(\lambda h) = \lambda \overline{\mathrm{T}}h$ [non-negative homogeneity]

*for arbitrary h, h_1, h_2 in $\mathcal{L}(\mathcal{X})$ and real $\lambda \geq 0$. The transformation $\underline{\mathrm{T}}$ of $\mathcal{L}(\mathcal{X})$, defined by $\underline{\mathrm{T}}f := -\overline{\mathrm{T}}(-f)$ for all $f \in \mathcal{L}(\mathcal{X})$, is then called its **conjugate lower transition operator**.*

Any upper transition operator $\overline{\mathrm{T}}$ automatically also satisfies the following interesting properties:

$\overline{\mathrm{T}}$4. $\overline{\mathrm{T}}(h + \mu I_{\mathcal{X}}) = \overline{\mathrm{T}}h + \mu I_{\mathcal{X}}$ [constant additivity]

$\overline{\mathrm{T}}$5. if $h_1 \leq h_2$ then $\overline{\mathrm{T}}h_1 \leq \overline{\mathrm{T}}h_2$ [monotonicity]

$\overline{\mathrm{T}}$6. if $h_n \to h$ (point-wise) then $\overline{\mathrm{T}}h_n \to \overline{\mathrm{T}}h$ [continuity]

$\overline{\mathrm{T}}$7. $\overline{\mathrm{T}}h \geq -\overline{\mathrm{T}}(-h) = \underline{\mathrm{T}}h$ [upper–lower consistency]

for arbitrary h, h_1, h_2, h_n in $\mathcal{L}(\mathcal{X})$ and real μ. Clearly, for any $n \in \mathbb{N}_{\geq 0}$, $\overline{\mathrm{T}}^n$ is an upper transition operator as well. As usual $\overline{\mathrm{T}}^0$ corresponds to the identity map.

3.3 Recursive nature of imprecise Markov chains

The upper previsions $\overline{\mathrm{P}}(\cdot|x_{1:n})$ and $\overline{\mathrm{P}}$ on $\mathcal{L}(\mathcal{X}^N)$ can be calculated easily using the Law of Iterated Expectation (Theorem 112), by backwards recursion, cf. (3.8) and (3.9).

Theorem 112 (Concatenation Formula): *For any $x_{1:n}$ in $\mathcal{X}^n$, $n = 1, 2, \ldots$, $N-1$, and for any real-valued map f on $\mathcal{X}^N$:*

$$\overline{\mathrm{P}}(f|x_{1:n}) = \overline{\mathbb{T}}_n \overline{\mathbb{T}}_{n+1} \ldots \overline{\mathbb{T}}_{N-1} f(x_{1:n}) \tag{3.12}$$

$$\overline{\mathrm{P}}(f) = \overline{\mathrm{Q}}_1(\overline{\mathbb{T}}_1 \overline{\mathbb{T}}_2 \ldots \overline{\mathbb{T}}_{N-1} f). \tag{3.13}$$

If we apply the joint upper prevision $\overline{\mathrm{P}}$ to maps h that depend only on the state X_n at time n, we get the **marginal upper previsions** $\overline{\mathrm{P}}_n(h) := \overline{\mathrm{P}}(h)$, and $\overline{\mathrm{P}}_n$ is a model for the uncertainty about the state X_n at time n. More generally, taking the Markov condition into account, we use the notation $\overline{\mathrm{P}}_{n|\ell}(h|x_\ell) := \overline{\mathrm{P}}_{|\ell}(h|x_\ell)$ for the upper prevision of $h(X_n)$, conditional on

$X_\ell = x_\ell$ with $1 \leq \ell < n$ and we see that $\overline{\mathrm{P}}_{n+1|n}(h|x_n) = \overline{\mathrm{Q}}_{n+1}(h|x_n) = \overline{\mathrm{T}}_n h(x_n)$. Such previsions can be found using simpler recursion formulae than Equations (3.12) and (3.13), as they are based on the simpler upper transition operators $\overline{\mathrm{T}}_k$.

Proposition 113: *For any real-valued map h on $\mathcal{X}$, and for any $1 \leq \ell < n \leq N$ and all x_ℓ in $\mathcal{X}$:*

$$\overline{\mathrm{P}}_{n|\ell}(h|x_\ell) = \overline{\mathrm{T}}_\ell \overline{\mathrm{T}}_{\ell+1} \dots \overline{\mathrm{T}}_{n-1} h(x_\ell), \tag{3.14}$$

$$\overline{\mathrm{P}}_n(h) = \overline{\mathrm{Q}}_1(\overline{\mathrm{T}}_1 \overline{\mathrm{T}}_2 \dots \overline{\mathrm{T}}_{n-1} h). \tag{3.15}$$

This offers a reason for formulating our theory in terms of real-valued maps rather than events: suppose we want to calculate the upper probability $\overline{\mathrm{P}}_n(A)$ that the state X_n at time n belongs to the set A. According to Equation (3.15), $\overline{\mathrm{P}}_n(A) = \overline{\mathrm{Q}}_1(\overline{\mathrm{T}}_1 \dots \overline{\mathrm{T}}_{n-1} I_A)$, and even if $\overline{\mathrm{T}}_{n-1} I_A$ can still be calculated using upper probabilities only, it will generally assume values other than 0 and 1, and therefore will generally not be the indicator of some event. Already after one step, i.e., in order to calculate $\overline{\mathrm{T}}_{n-2}\overline{\mathrm{T}}_{n-1} I_A$, we need to leave the ambit of events, and turn to the more general real-valued maps; even if we only want to calculate upper probabilities after n steps.

From here onwards, we assume that the imprecise Markov chain is **stationary**: $\overline{\mathrm{T}}_1 = \overline{\mathrm{T}}_2 = \dots = \overline{\mathrm{T}}$. For stationary imprecise Markov chains, Proposition 113 simplifies.

Proposition 114: *For any real-valued map h on $\mathcal{X}$, and for any $1 \leq \ell < n \leq N$ and all x_ℓ in $\mathcal{X}$:*

$$\overline{\mathrm{P}}_{n|\ell}(h|x_\ell) = \overline{\mathrm{T}}^{n-\ell} h(x_\ell), \tag{3.16}$$

$$\overline{\mathrm{P}}_n(h) = \overline{\mathrm{Q}}_1(\overline{\mathrm{T}}^{n-1} h). \tag{3.17}$$

3.4 Sensitivity interpretation

The classical treatment of Markov chains rests on the assumption that the initial and transition probabilities are precisely known. If this is not the case, then it seems necessary to perform some kind of sensitivity analysis, in order to find out to what extent any conclusion we might reach using such a treatment, depends on the actual values of these probabilities.

To see what is the link between upper transition operators and sets of precise finite-state and discrete-time Markov chains, consider the coherent upper prevision $\overline{\mathrm{T}}h(x) \coloneqq \overline{\mathrm{Q}}(h|x)$. Because $\overline{\mathrm{Q}}(\cdot|x)$ is monotone ($\overline{\mathrm{T}}5_{130}$), constant additive ($\overline{\mathrm{T}}4_{130}$), convex ($\overline{\mathrm{T}}2_{130}$+$\overline{\mathrm{T}}3_{130}$) and non-negatively homogeneous ($\overline{\mathrm{T}}3_{130}$), it follows from Legendre-Fenchel duality (and the Lower Envelope Theorem [62, §2.6.3] and to a lesser extent Theorem 18_{41}), that $\overline{\mathrm{Q}}(h|x)$ can be written as (see [1])[5]

$$\overline{\mathrm{Q}}(h|x) = \max\{p \cdot h \colon p \in \mathcal{P}_x\},$$

where $\mathcal{P}_x$ is a compact convex set of probability mass functions, also known as a **credal set**. The upper transition operator $\overline{\mathrm{T}}h$ can now be seen as the Cartesian product (or a vector) of the upper previsions over all states. If given a prior upper prevision $\overline{\mathrm{Q}}_1$ corresponding to a credal set $\mathcal{P}_1$:

$$\overline{\mathrm{Q}}_1(h) = \max\{p_1 \cdot h \colon p_1 \in \mathcal{P}_1\}, \tag{3.18}$$

then it follows almost immediately that

$$\overline{\mathrm{Q}}_1(\overline{\mathrm{T}}h) = \max\{p_1 \cdot M \cdot h \colon p_1 \in \mathcal{P}_1 \text{ and } M \in \mathcal{T}\},$$

where

$$\mathcal{T} \coloneqq \left\{M \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|} \colon (\forall x \in \mathcal{X})(M_{x,\cdot} \in \mathcal{P}_x)\right\}. \tag{3.19}$$

Here, any $M \in \mathcal{T}$ is a stochastic matrix where the x-th row, $M_{x,\cdot}$, is a probability distribution over the states at a time $k+1$, conditional on the chain being in state x at time k. Therefore, we can interpret M as a transition matrix of a finite-state and discrete-time Markov chain. When considering iterations of the map, then we see that

$$\overline{\mathrm{Q}}_1(\overline{\mathrm{T}}^k h) = \max\left\{p_1 \cdot M^{(1)} \cdot \ldots \cdot M^{(k)} \cdot h \colon p_1 \in \mathcal{P}_1 \text{ and } M^{(j)} \in \mathcal{T}\right\}. \tag{3.20}$$

Generally speaking, therefore, an upper transition operator effects robust inference for a set of not necessarily stationary Markov chains **whenever we are investigating marginal gambles**. In general, however, the sensitivity interpretation and our definition of a Markov chain are not the same. The difference lies in the type of independence that is assumed in the Markov Condition. Whereas we assume epistemic irrelevance, **strong independence** is assumed when using the sensitivity interpretation [8].

[5] Here we use the inner product notation $p \cdot h \coloneqq \sum_{x\in\mathcal{X}} p(x)h(x)$.

3.4.1 When interpretation does not matter

The question arises whether the interpretation of the Markov Condition really matters. The next example shows that in general, it actually does.

▷ **Example 115:** Let $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3 = \{a, b\}$ and $\overline{\mathrm{Q}}_1(h) = \mathrm{Q}_1(fh) = \frac{h(a)+h(b)}{2}$. Then we know from Equation (3.18) that $\mathcal{P}_1 = \left\{\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\end{pmatrix}\right\}$. The upper transition operator is given by a vacuous model, independent of the state: $\overline{\mathrm{T}} := I_{\mathcal{X}} \max$. Therefore

$$\mathcal{T} := \left\{M \in \mathbb{R}^{2\times 2} \colon M_{a,\cdot} \in \mathcal{P}_a \text{ and } M_{b,\cdot} \in \mathcal{P}_b\right\}.$$

with $\mathcal{P}_a = \mathcal{P}_b = \left\{\begin{pmatrix}p & 1-p\end{pmatrix} \colon 0 \le p \le 1\right\}$. The gamble $f \in \mathcal{L}(\mathcal{X}_{1:3})$ of interest is $f = I_{(a,a,a)} + I_{(b,a,b)}$.

Under the strong independence interpretation, the joint credal set $\mathcal{P}_{1:3}$ is given by

$$\mathcal{P}_{1:3} := \mathcal{P}_1\mathcal{T}^2 = \left\{p_1 \cdot M^{(1)} \cdot M^{(2)} \cdot h \colon p_1 \in \mathcal{P}_1 \text{ and } M^{(j)} \in \mathcal{T}\right\}$$

and we get that

$$\begin{aligned}
\overline{\mathrm{P}}(f) &= \max\left\{\sum_{x\in\mathcal{X}_{1:3}} p(x)f(x) \colon p \in \mathcal{P}_{1:3}\right\},\\
&= \max\{p(a,a,a) + p(b,a,b) \colon p \in \mathcal{P}_{1:3}\},\\
&= \max\{q_1(a)q_2(a|a)q_3(a|a) + q_1(b)q_2(a|b)q_3(b|a) \colon\\
&\qquad q_1 \in \mathcal{P}_1, q_2(\cdot|a) \in \mathcal{P}_a, q_2(\cdot|b) \in \mathcal{P}_b, q_3(\cdot|a) \in \mathcal{P}_a, q_3(\cdot|b) \in \mathcal{P}_b\}\\
&= \max\left\{\frac{1}{2}q_2(a|a)q_3(a|a) + \frac{1}{2}q_2(a|b)[1 - q_3(a|a)] \colon\right.\\
&\qquad \left. q_2(a|a) \in [0,1], q_2(a|b) \in [0,1], q_3(a|a) \in [0,1]\right\}\\
&= \frac{1}{2}.
\end{aligned}$$

Under the epistemic irrelevance interpretation, we get from Theorem 112$_{130}$ that

$$\begin{aligned}
\overline{\mathrm{P}}(f) &= \overline{\mathrm{Q}}_1\left(\sum_{x_1\in\mathcal{X}} I_{\{x_1\}}\overline{\mathrm{Q}}_2\left(\sum_{x_2\in\mathcal{X}} I_{\{x_2\}}\overline{\mathrm{Q}}_3(f(x_1,x_2,\cdot)|x_2)\middle| x_1\right)\right)\\
&= \frac{1}{2}\overline{\mathrm{Q}}_2\left(I_{\{a\}}\overline{\mathrm{Q}}_3(f(a,a,\cdot)|a) + I_{\{b\}}\overline{\mathrm{Q}}_3(f(a,b,\cdot)|b)\middle| a\right)\\
&\qquad + \frac{1}{2}\overline{\mathrm{Q}}_2\left(I_{\{a\}}\overline{\mathrm{Q}}_3(f(b,a,\cdot)|a) + I_{\{b\}}\overline{\mathrm{Q}}_3(f(b,b,\cdot)|b)\middle| b\right)\\
&= \frac{1}{2}\overline{\mathrm{Q}}_2\left(I_{\{a\}}\overline{\mathrm{Q}}_3(I_{\{a\}}|a)\middle| a\right) + \frac{1}{2}\overline{\mathrm{Q}}_2\left(I_{\{a\}}\overline{\mathrm{Q}}_3(I_{\{b\}}|a)\middle| b\right) \qquad (3.21)\\
&= \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1 = 1.
\end{aligned}$$

Rewriting Equation (3.21) gives more insight about where the difference comes from

$$\begin{aligned}
\overline{\mathrm{P}}(f) &= \max\Big\{ q_1(a)q_2(a|a)q_3(a|a) + q_1(b)q_2'(a|b)q_3'(b|a) : \\
&\qquad q_1 \in \mathcal{P}_1, q_2(\cdot|a) \in \mathcal{P}_a, q_2'(\cdot|b) \in \mathcal{P}_b, q_3(\cdot|a) \in \mathcal{P}_a, q_3'(\cdot|a) \in \mathcal{P}_a, \Big\} \\
&= \max\Big\{ \frac{1}{2} q_2(a|a)q_3(a|a) + \frac{1}{2} q_2'(a|b)q_3'(b|a) : \\
&\qquad q_2(a|a) \in [0,1], q_2'(a|b) \in [0,1], q_3(a|a) \in [0,1], q_3'(b|a) \in [0,1] \Big\} \\
&= \frac{1}{2} \cdot 1 \cdot 1 + \frac{1}{2} \cdot 1 \cdot 1 = 1.
\end{aligned}$$

The inequality still holds if the transition operator is assumed to be a linear-vacuous mixture (not purely linear).

The example shows that the interpretation is relevant. It also shows that the sensitivity interpretation result can be obtained from the epistemically irrelevant one by adding extra constraints. The extra constraints are that, at any given time point, the optimisation is allowed to choose only one transition matrix. Hence, the imprecise Markov chain is a set of (non-stationary) precise Markov models under the strong independence interpretation. This is not the case in the epistemic irrelevance case where the choice of transition matrices may depend on the complete history.

The discrepancy between the two interpretations is most unfortunate. However, if the gamble depends only on one state, then it follows immediately from Equation (3.20) that the interpretation is irrelevant. This is so, because the local optimisation done in the epistemic irrelevance case takes the same transition matrix, which will of course be the one that was used in the strong independence case.

Theorem 116: *Let $\{X_n : n \in \mathbb{N}_{>0}\}$ be a stationary imprecise Markov chain and let the upper prevision functional $\overline{\mathrm{Q}}_1$ represent the beliefs about the initial state X_1 and $\overline{\mathrm{T}}$ be the upper transition operator. Then the upper prevision of a gamble h depending only on one state X_n is given by*

$$\overline{\mathrm{P}}_n(h) = \overline{\mathrm{Q}}_1(\overline{\mathrm{T}}^{n-1} h), \tag{3.22}$$

independent of the assumed Markov Condition.

3.5 Graph-theoretic analysis of upper transition operators

Let us introduce the notation [14]:

$$\overline{P}^{n}_{xy} := \overline{\mathrm{T}}^{n} I_{\{y\}}(x) \text{ for } n \in \mathbb{N}_{\geq 0}. \tag{3.23}$$

$\overline{P}^{n}_{xy}$ can be interpreted as an upper probability of going from state x to state y in n steps. For $n = 0$, $\overline{P}^{0}_{xy} = I_{\{y\}}(x)$, and for $n = 1$ we also use the simpler notation $\overline{P}_{xy}$ instead of $\overline{P}^{1}_{xy}$. The following inequality is of crucial importance for what follows. It is an immediate generalisation of a similar equality (Chapman–Kolmogorov) involving (precise) probabilities in (precise) Markov chains.

Proposition 117: *For all x, y and z in $\mathcal{X}$, and for all m and n in $\mathbb{N}_{\geq 0}$,*

$$\overline{P}^{n+m}_{xy} \geq \overline{P}^{n}_{xz}\overline{P}^{m}_{zy}. \tag{3.24}$$

Proof: Since $\overline{P}^{m}_{uy} = \overline{\mathrm{T}}^{m} I_{\{y\}}(u) \geq 0$ for all $u \in \mathcal{X}$, we have that

$$\overline{\mathrm{T}}^{m} I_{\{y\}} = \sum_{u\in\mathcal{X}} \overline{\mathrm{T}}^{m} I_{\{y\}}(u) I_{\{u\}} \geq \overline{\mathrm{T}}^{m} I_{\{y\}}(z) I_{\{z\}} \tag{3.25}$$

for all $z \in \mathcal{X}$. If we now apply the upper transition operator $\overline{\mathrm{T}}$ n times to both sides of this inequality, and repeatedly invoke its monotonicity [$\overline{\mathrm{T}}5_{130}$] and non-negative homogeneity [$\overline{\mathrm{T}}3_{130}$], we find that $\overline{\mathrm{T}}^{n+m} I_{\{y\}} \geq \overline{\mathrm{T}}^{m} I_{\{y\}}(z) \overline{\mathrm{T}}^{n} I_{\{z\}}$ and hence indeed $\overline{\mathrm{T}}^{n+m} I_{\{y\}}(x) \geq \overline{\mathrm{T}}^{n} I_{\{z\}}(x) \overline{\mathrm{T}}^{m} I_{\{y\}}(z)$. □

Using the $\overline{P}^{n}_{xy}$, we can define a ternary relation $\cdot \overset{\cdot}{\to} \cdot$ on $\mathcal{X} \times \mathcal{X} \times \mathbb{N}_{\geq 0}$ as follows: for any x and y in $\mathcal{X}$ and any $n \in \mathbb{N}_{\geq 0}$, we say that y **is accessible from** x **in** n **steps**, and we write $x \overset{n}{\to} y$, if $\overline{P}^{n}_{xy} > 0$:

$$x \overset{n}{\to} y \Leftrightarrow \overline{P}^{n}_{xy} > 0 \Leftrightarrow \overline{\mathrm{T}}^{n} I_{\{y\}}(x) > 0. \tag{3.26}$$

Proposition 118: *The ternary relation $\cdot \overset{\cdot}{\to} \cdot$ is an* ***accessibility relation****, meaning that:*

$$(\forall x, y \in \mathcal{X})(x \overset{0}{\to} y \Leftrightarrow x = y); \tag{C1}$$

$$(\forall x, y \in \mathcal{X})(\forall m, n \in \mathbb{N}_{\geq 0})\,(x \overset{n+m}{\to} y \Leftrightarrow (\exists z \in \mathcal{X})(x \overset{n}{\to} z \text{ and } z \overset{m}{\to} y)); \tag{C2}$$

$$(\forall x \in \mathcal{X})(\forall n \in \mathbb{N}_{\geq 0})(\exists y \in \mathcal{X})\, x \overset{n}{\to} y. \tag{C3}$$

Proof:

C1 This property follows at once from Equation (3.23).

C2 The implication $(\exists z \in \mathcal{X})(x \overset{n}{\rightarrow} z \text{ and } z \overset{m}{\rightarrow} y) \Rightarrow x \overset{n+m}{\rightarrow} y$ follows immediately from Equation (3.24).

Assume that $x \overset{n+m}{\rightarrow} y$, then we know that

$$0 < \overline{\mathrm{T}}^{n+m} I_{\{y\}}(x) = \overline{\mathrm{T}}^{n}\left(\overline{\mathrm{T}}^{m} I_{\{y\}}\right)(x) = \overline{\mathrm{T}}^{n}\left(\sum_{z\in\mathcal{X}} I_{\{z\}} \overline{\mathrm{T}}^{m} I_{\{y\}}(z)\right)(x)$$

and using the sub-additivity property $\overline{\mathrm{T}}2_{130}$ we get that

$$0 < \sum_{z\in\mathcal{X}} \overline{\mathrm{T}}^{n} I_{\{z\}}(x) \overline{\mathrm{T}}^{m} I_{\{y\}}(z)$$

which can—knowing that both $\overline{\mathrm{T}}^{n} I_{\{z\}}(x)$ and $\overline{\mathrm{T}}^{m} I_{\{y\}}(z)$ are nonnegative by property $\overline{\mathrm{T}}1_{130}$— only be true if there is some $z \in \mathcal{X}$ such that $\overline{\mathrm{T}}^{n} I_{\{z\}}(x) > 0$ and $\overline{\mathrm{T}}^{m} I_{\{y\}}(z) > 0$ and therefore $x \overset{n+m}{\rightarrow} y \Rightarrow x \overset{n}{\rightarrow} z \wedge z \overset{m}{\rightarrow} y$.

C3 Consider any $x \in \mathcal{X}$ and $n \in \mathbb{N}_{\geq 0}$. Assume *ex absurdo* that $\overline{\mathrm{T}}^{n} I_{\{y\}}(x) = 0$ for all $y \in \mathcal{X}$. Since $\overline{\mathrm{T}}^{n}$ is an upper transition operator, it follows from $\overline{\mathrm{T}}1_{130}$ and $\overline{\mathrm{T}}2_{130}$ that $1 \leq \overline{\mathrm{T}}^{n} I_{\mathcal{X}} \leq \sum_{y\in\mathcal{X}} \overline{\mathrm{T}}^{n} I_{\{y\}}$, whence $1 \leq 0$, a contradiction. □

The last condition C3 was assumed implicitly by Kemeny and Snell [35, § 1.4]. We have made it explicit here as it is exactly this requirement that forces us to work with upper instead of lower transition operators. If we take the lower transition operator $\underline{\mathrm{T}} := I_{\mathcal{X}} \min$, then $\underline{\mathrm{T}} I_{\{x\}} = 0$ and C3 does not hold.

Kemeny and Snell did not assume an equivalence in condition C2, they only required that

$$(\forall x, y \in \mathcal{X})(\forall m, n \in \mathbb{N}_{\geq 0})\left((\exists z \in \mathcal{X})(x \overset{n}{\rightarrow} z \text{ and } z \overset{m}{\rightarrow} y) \Rightarrow x \overset{n+m}{\rightarrow} y\right). \quad \text{(C2')}$$

All qualitative statements that can be made for the accessibility relation can also be made with the alternative requirement C2'. However, the computation of accessibility properties—like periodicity, regularity, ergodicity—becomes very hard because the link with the accessibility graph will be lost. This is why we assume the stronger condition C2 as it will allow us to deduce all interesting results from this simple graph: the accessibility graph.[6]

Definition 119: *The* ***accessibility graph*** $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ *of an accessibility relation* $\cdot \overset{\cdot}{\rightarrow} \cdot : \mathcal{X} \times \mathcal{X} \times \mathbb{N}_{\geq 0} \to \{0,1\}$ *is the directed graph with vertices* $\mathcal{X}$ *and an edge from* $x \in \mathcal{X}$ *to* $y \in \mathcal{X}$ *if and only if* $x \overset{1}{\rightarrow} y$.

[6] For more information about the accessibility graph and the graph terminology used in this chapter (e.g. reachability relation, strongly connected component, ...), we refer to Appendix A_{192}.

It is clear that a state y is accessible from x in n steps if and only if there is an (x, y)-path of length n in the accessibility graph $\overline{\mathcal{G}}(\overline{\mathrm{T}})$. This means that all results that can be deduced from the accessibility relation, can also be found by studying the accessibility graph.

3.5.1 State classification

We call **any** ternary relation $\cdot \overset{\cdot}{\rightarrow} \cdot$ that satisfies $\mathrm{C1}_{135}$–$\mathrm{C3}_{135}$ an **accessibility relation** on the finite set $\mathcal{X}$. For any such (abstract) accessibility relation, we can draw all the following conclusions. The present discussion is a formalisation of the more intuitive discussion by Kemeny and Snell [35, § 1.4], under slightly more restrictive conditions. We refer to Figure 3.3 for a graphical representation of many of the notions discussed below.

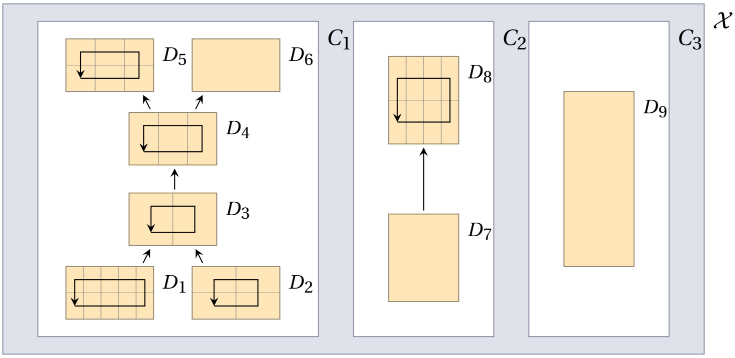


Figure 3.3: Three increasingly finer partitions of the state set $\mathcal{X}$ for a particular stationary (im)precise Markov chain, or more generally, for an accessibility relation $\cdot \overset{\cdot}{\rightarrow} \cdot$. No transition between states in different closed atoms C_k is possible, and closed atoms can be seen as separate (im)precise Markov chains. The equivalence classes D_k for the communication relation are partially ordered by the relation $\rightarrow$, whose (Hasse) diagram is represented by the upward arrows. Maximal classes are D_5, D_6, D_8 and D_9, the other classes are non-maximal. Each communication class D_k is further subdivided in d_{D_k} cyclic classes, through which the system can move in a cyclic fashion, e.g., $d_{D_2} = 4$. For a non-maximal class, it is also possible to move to descendant classes, until finally a maximal class is reached. There are 3 aperiodic classes: D_6, D_7 and $D_9 = C_3$. The (im)precise Markov sub-chain C_3 is also irreducible, and therefore regular.

Consider any two states x and y in $\mathcal{X}$. Then y is **accessible from** x, which we denote as $x \rightarrow y$, if there is some $n \in \mathbb{N}_{\geq 0}$ such that $x \overset{n}{\rightarrow} y$. If x and y are accessible from one another, then we say that x and y **communicate**, which we denote as $x \leftrightarrow y$.

It follows at once from C1135 and C2135 that the binary relation $\to$ on $\mathcal{X}$ is a preorder, i.e., is reflexive and transitive. The binary relation $\leftrightarrow$ on $\mathcal{X}$ is the associated equivalence relation. This communication relation $\leftrightarrow$ partitions the state set $\mathcal{X}$ into equivalence classes D of states that are accessible from one another, called **communication classes**. The preorder $\to$ induces a partial order on this partition, also denoted by $\to$.

Maximal (or undominated) states with respect to the preorder $\to$ are states x such that $(\forall y \in \mathcal{X})(x \to y \Rightarrow y \to x)$. This means that a maximal state only has access to other maximal states in the same communication class, and to no other states. They, and the communication classes they belong to, are called **maximal**. The other states, and the communication classes they belong to, are called **non-maximal** or **dominated**.

In terms of the accessibility graph $\overline{\mathcal{G}}(\overline{\mathrm{T}})$, the reachability relation (see Appendix A192 for concepts and related terminology) is the counterpart of the accessibility relation: If x has access to y, then either $x = y$, or there is a (x, y)-path in $\overline{\mathcal{G}}(\overline{\mathrm{T}})$.

Proposition 120: *The communication classes of an upper transition operator $\overline{\mathrm{T}}$, correspond exactly to the* ***strongly connected components*** *of $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ and the* ***maximal communication classes*** *are the undominated strong components of $\overline{\mathcal{G}}(\overline{\mathrm{T}})$.*

3.5.2 Periodicity

Consider, for any x and y in $\mathcal{X}$, the set

$$N_{xy} := \{n \in \mathbb{N}_{\geq 0} : x \overset{n}{\to} y\} \tag{3.27}$$

of those numbers of steps after which y is accessible from x. Observe that N_{xy} is non-empty if and only if $x \to y$.

N_{xx} always contains at least 0. If N_{xx} contains no other elements, then x constitutes a communication class by itself. In terms of an upper transition operator $\overline{\mathrm{T}}$ this requires that $\overline{\mathrm{T}}^n I_{\{x\}}(x) = 0$ for all $n > 0$. If a system enters such a state x, it must immediately leave it and can never come back to it. We call any such state, and the communication class it constitutes, **volatile**. We infer from condition C3135 that no maximal state can be volatile.

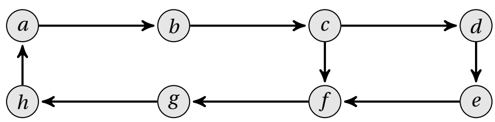
Definition 121: *We call the **period** d_x of a state x the greatest common divisor of the non-empty set N_{xx}, i.e.,*

$$d_x := \gcd\left\{n \in \mathbb{N}_{\geq 0} \colon x \xrightarrow{n} x\right\}.$$

This means that the lengths of all (x,x)-paths in $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ must be multiples of d_x. It is not necessary though, that there is some (x,x)-path for every multiple of d_x.

▷ **Example 122:** For the accessibility graph given in the figure, returns to state a can only occur for

$$N_{aa} = \left\{\alpha 6 + \beta 8 \colon \alpha \in \mathbb{N}_{\geq 0}, \beta \in \mathbb{N}_{\geq 0}\right\}.$$



Therefore, the period $d_a = \gcd N_{aa} = 2$. Only paths from a to a exist that have a length that is a multiple of 2. There is no (a,a)-path of length 2 or 4.

The complete state space is a communication class as the accessibility graph is strongly connected. The period for every state is 2 (see also Proposition 124).

A state is volatile if and only if its period is infinite. If $d_x < +\infty$, then because, by C2$_{135}$, the set N_{xx} is closed under addition, the basic number-theoretic result of Theorem 123 tells us that N_{xx} is, up to perhaps a finite number of elements, equal to the set of all multiples of d_x:

$$(\exists n \in \mathbb{N}_{\geq 0})(\forall k \geq n)\, k d_x \in N_{xx}. \tag{3.28}$$

Theorem 123 ([35, Theorem 1.4.1]): *A non-empty set of natural numbers that is closed under addition contains all but a finite number of multiples of its greatest common divisor.*

Now consider a communication class D, and any two states x and y in that class. Then it is not difficult to show that they have the same period.

Proposition 124: *For any two elements x and y of $\mathcal{X}$: $x \leftrightarrow y \Rightarrow d_x = d_y$.*

Proof: We assume without loss of generality that $x \neq y$. Suppose $x \leftrightarrow y$. Then $N_{xy} \neq \emptyset$ and $N_{yx} \neq \emptyset$. Fix $n_{xy} \in N_{xy}$ and $n_{yx} \in N_{yx}$. Then $n_{xy} + n_{yx} \in N_{xx} \cap N_{yy}$. Since

$n_{xy} + n_{yx} > 0$ this implies that both d_x and d_y are finite. Hence there are r_x and r_y in $\mathbb{N}_{\geq 0}$ such that

$$r_x d_x = n_{xy} + n_{yx} = r_y d_y. \tag{3.29}$$

On the other hand, we infer from (3.28) that for sufficiently large $k \in \mathbb{N}_{\geq 0}$ both $n_{xy} + n_{yx} + k d_y$ and $n_{xy} + n_{yx} + (k+1) d_y$ are elements of N_{xx}. Using (3.29), this implies that d_x divides both $(r_y + k) d_y$ and $(r_y + k + 1) d_y$. Hence d_x divides d_y. A completely similar argument shows that d_y divides d_x. □

We denote by d_D the common period of all elements of the communication class D. If $d_D = 1$, then we call the class D, and all its states, **aperiodic**.

The analysis can still be taken further in interesting ways.

Proposition 125: *Consider arbitrary x and y in some non-volatile communication class D. Then there is some $0 \leq t_{xy} < d_D$ such that $n \in N_{xy}$ implies $n \equiv t_{xy} \pmod{d_D}$, i.e., n and t_{xy} are equal up to some multiple of d_D. Moreover,*

$$(\exists n \in \mathbb{N}_{\geq 0})(\forall k \geq n)\, t_{xy} + k d_D \in N_{xy}, \tag{3.30}$$

and therefore N_{xy} equals $\{t_{xy} + k d_D \colon k \in \mathbb{N}_{\geq 0}\}$ up to perhaps a finite number of elements. Finally,

$$(\forall x, y, z \in D)(t_{xy} + t_{yz} \equiv t_{xz} \pmod{d_D}). \tag{3.31}$$

Proof: Since $N_{xy} \neq \emptyset$ and $N_{yx} \neq \emptyset$ we can consider arbitrary n_{xy} and m_{xy} in N_{xy} and arbitrary n_{yx} in N_{yx}. Since both $n_{xy} + n_{yx}$ and $m_{xy} + n_{yx}$ belong to N_{xx}, we see that $n_{xy} \equiv m_{xy} \pmod{d_D}$, so all elements of N_{xy} have the same remainder after division by d_D. If we call this common remainder t_{xy}, then obviously $0 \leq t_{xy} < d_D$ and $n_{xy} \equiv t_{xy} \pmod{d_D}$ for all $n_{xy} \in N_{xy}$.

To prove (3.30), consider any $n_{xy} \in t_{xy}$. We have just proved that there is some $n' \in \mathbb{N}_{\geq 0}$ such that $n_{xy} = t_{xy} + n' d_D$. And we know from (3.28) that there is some $n'' \in \mathbb{N}_{\geq 0}$ such that for all $k \geq n''$, $k d_D \in N_{xx}$ and therefore $t_{xy} + (n' + k) d_D \in N_{xy}$.

To prove (3.31), consider $n_{xy} \in N_{xy}$ and $n_{yz} \in N_{yz}$, then on the one hand $n_{xy} + n_{yz} \in N_{xz}$ and therefore $n_{xy} + n_{yz} \equiv t_{xz} \pmod{d_C}$. On the other hand, if follows from $n_{xy} \equiv t_{xy} \pmod{d_C}$ and $n_{yz} \equiv t_{yz} \pmod{d_C}$ that $n_{xy} + n_{yz} \equiv t_{xy} + t_{yz} \pmod{d_C}$. □

It follows that for any x, y and z in a non-volatile communication class D, $t_{xx} = 0$ and $t_{xy} + t_{yz} \equiv t_{xz} \pmod{d_D}$, and therefore $t_{yz} = 0$ if and only if $t_{xy} = t_{xz}$ for some $x \in D$. This implies that '$t_{yz} = 0$' determines an equivalence relation on this communication class D, which further partitions it into d_D

subsets, called **cyclic classes**. In such a cyclic class, all states y give the same value to t_{xy}, for any given x in D. Within D, the system moves from cyclic class to cyclic class, in a definite ordered cycle of length d_D. If D is non-maximal, then in each of its cyclic classes it is possible that, rather than moving to the next cyclic class, the system moves on to (a state in) another communication class D', that is a successor to D for the partial order $\rightarrow$.

Definition 126: *A maximal aperiodic communication class is called* ***regular****. If there is only one communication class, then $\mathcal{X}$ is called* ***irreducible****. If $\mathcal{X}$ is irreducible and aperiodic, $\mathcal{X}$ itself is also called* ***regular****.*

The following characterisations of regularity are now immediate (observe that $d_D = 1$ and $t_{xy} = 0$ for an aperiodic class).

Proposition 127: *A maximal communication class D is regular if and only if*

$$(\exists n \in \mathbb{N}_{\geq 0})(\forall k \geq n)(\forall x, y \in D)(x \stackrel{k}{\rightarrow} y).$$

$\mathcal{X}$ is regular if and only if

$$(\exists n \in \mathbb{N}_{\geq 0})(\forall k \geq n)(\forall x, y \in \mathcal{X})(x \stackrel{k}{\rightarrow} y).$$

Let us define the set of all (simple) cycles $\mathcal{C}_x$ from a state x to itself.

$$\mathcal{C}_x := \left\{x \stackrel{1}{\rightarrow} x_1 \stackrel{1}{\rightarrow} x_2 \stackrel{1}{\rightarrow} \dots \stackrel{1}{\rightarrow} x \colon (x_i \neq x) \text{ and } (i \neq j \Rightarrow x_i \neq x_j)\right\}.$$

The length of a path $c := x_0 \stackrel{1}{\rightarrow} x_1 \stackrel{1}{\rightarrow} x_2 \stackrel{1}{\rightarrow} \dots \stackrel{1}{\rightarrow} x_{n-1} \stackrel{1}{\rightarrow} x_n$ is by definition $\text{length}(c) = n$.

Proposition 128: *The period d_D of a communication class D is equal to*

$$d_D = \gcd\{\text{length}(c) \colon c \in \mathcal{C}_x\},$$

where x is any element of D.

Proof: Remark that any "path"[7] from x to x has to be a concatenation of elements of $\mathcal{C}_x$. We infer then from Definition 121$_{139}$ and from Proposition 124$_{139}$ that

$$d_C = \gcd\left\{\sum_{c \in \mathcal{C}_c} a_c \,\text{length}(c) \colon a_c \in \mathbb{N}_{\geq 0}\right\} = \gcd\{\text{length}(c) \colon c \in \mathcal{C}x\}. \qquad \square$$

[7]Here we mean with "path" a sequence of states where consecutive states in the sequence are accessible in one step. Remark that states can be repeated in this "path" whence it is not a path in the graph-theoretic sense of Appendix A$_{192}$

Remark that $x \xrightarrow{1} y$ if and only if there is an arc (x, y) in $\overline{\mathcal{G}}(\overline{\mathrm{T}})$. All this means that the search for the period of a communication class C can be reduced to finding the period of the corresponding strongly connected component $\overline{\mathcal{G}}(\overline{\mathrm{T}})\big|_C$.[8] This problem is closely related to a breadth-first search in $\overline{\mathcal{G}}(\overline{\mathrm{T}})\big|_C$ as explained in [25].

3.5.3 Adding more detail to this picture

We now return from this abstract discussion to the specific accessibility relation $\rightarrow$ associated with an upper transition operator $\overline{\mathrm{T}}$. Using $\overline{\mathrm{T}}$, we will be able to add some detail to the sketch made using the relation $\rightarrow$ only.

We begin by looking at closed sets of states. A set of states C is called **closed** if no state outside C is accessible from any state within the set of states $C : (\forall x \in C)(\forall y \in C^c)(x \not\rightarrow y)$. In particular, any maximal communication class is a closed set of states. Of course, $\mathcal{X}$ itself is also closed. And so is any union of closed sets.

Proposition 129: *Consider a stationary imprecise Markov chain with upper transition operator $\overline{\mathrm{T}}$. Let $\mathcal{C}$ be a partition of the state set $\mathcal{X}$ into closed sets and let $C \in \mathcal{C}$.*[9] *Then*

1. $\overline{\mathrm{T}}(hI_B)(x) = 0$ *for all $h \in \mathcal{L}(\mathcal{X})$, all $x \in C$ and all $B \subseteq C^c$;*

2. $\overline{\mathrm{T}}h(x) = \overline{\mathrm{T}}(hI_C)(x)$ *for all $h \in \mathcal{L}(\mathcal{X})$ and all $x \in C$;*

3. $\overline{\mathrm{T}}h = \sum_{C\in\mathcal{C}} \overline{\mathrm{T}}(I_C h) = \sum_{C\in\mathcal{C}} I_C \overline{\mathrm{T}}(I_C h)$ *for all $h \in \mathcal{L}(\mathcal{X})$.*

Proof: First, fix x in C and $B \subseteq C^c$. Since the set C is closed, it follows that for any y in B, $\overline{\mathrm{T}}I_{\{y\}}(x) = \overline{P}_{xy} = 0$. Using $\overline{\mathrm{T}}2_{130}$ and $\overline{\mathrm{T}}5_{130}$, we see that therefore $\overline{\mathrm{T}}I_B(x) = 0$. But since $-I_B \max|h| \leq hI_B \leq I_B \max|h|$, we find that on the one hand, using $\overline{\mathrm{T}}5_{130}$ and $\overline{\mathrm{T}}3_{130}$, $\overline{\mathrm{T}}(hI_B) \leq (\max|h|)\overline{\mathrm{T}}I_B$. On the other hand, using $\overline{\mathrm{T}}5_{130}$, $\overline{\mathrm{T}}7_{130}$ and $\overline{\mathrm{T}}3_{130}$, we get $-(\max|h|)\overline{\mathrm{T}}I_B \leq \overline{\mathrm{T}}(hI_B)$. Since $\overline{\mathrm{T}}I_B(x) = 0$, we see that $\overline{\mathrm{T}}(hI_B)(x) = 0$, which proves the first statement.

We infer from the first statement that both $\overline{\mathrm{T}}(hI_{C^c})(x) = 0$ and $\overline{\mathrm{T}}(-hI_{C^c})(x) = 0$. Since $h = hI_C + hI_{C^c}$, it follows from $\overline{\mathrm{T}}2_{130}$ that $\overline{\mathrm{T}}(hI_C) - \overline{\mathrm{T}}(-hI_{C^c}) \leq \overline{\mathrm{T}}h \leq \overline{\mathrm{T}}(hI_C) + \overline{\mathrm{T}}(hI_{C^c})$ and therefore $\overline{\mathrm{T}}h(x) = \overline{\mathrm{T}}(hI_C)(x)$, which proves the second statement.

[8] Remember that $\overline{\mathcal{G}}(\overline{\mathrm{T}})\big|_C$ is the subgraph of $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ induced by C, see Appendix A$_{192}$.

[9] Remark that a partition of closed sets must always be a coarsening of the partition of communication classes. The set of all states $\mathcal{X}$ is always a closed set and forms the coarsest partition.

Consider any $C \in \mathcal{C}$ and any $y \in C^c$. Then y belongs to some closed set $C' \in \mathcal{C} \setminus \{C\}$, so we infer from the first statement that $\mathrm{T}(hI_C)(y) = 0$. This, together with the second statement, leads at once to the third statement. □

If we consider a closed set of states C, then we can define an operator $\overline{\mathrm{T}}_C : \mathcal{L}(C) \to \mathcal{L}(C)$ as follows. Consider any real-valued map g on C, and any extension h of g to $\mathcal{X}$. Then it follows from Proposition 129.2 that $\overline{\mathrm{T}}h = \overline{\mathrm{T}}(hI_C)$ only depends on g, and not on how g is extended to h outside C. It therefore makes sense to define $\overline{\mathrm{T}}_C g$ as the restriction $\overline{\mathrm{T}}h\big|_C$ of $\overline{\mathrm{T}}h$ to C. It is very easy to see that $\overline{\mathrm{T}}_C$ satisfies $\overline{\mathrm{T}}1_{130}$–$\overline{\mathrm{T}}3_{130}$, and is therefore an upper transition operator on $\mathcal{L}(C)$.

It also follows from Proposition 129.3, that if $\mathcal{C}$ is a collection of mutually disjoint closed subsets of the state set $\mathcal{X}$, then $\overline{\mathrm{T}}^n_{\bigcup \mathcal{C}} = \sum_{C \in \mathcal{C}} I_C \overline{\mathrm{T}}^n_C \circ |_C$ for all $n \geq 0$: the dynamics associated with $\overline{\mathrm{T}}_{\bigcup \mathcal{C}}$ on the closed subset $\bigcup \mathcal{C}$ of $\mathcal{X}$ can be seen as being subdivided into isolated and independent dynamics associated with $\overline{\mathrm{T}}_C$ on state sets $C \in \mathcal{C}$.

3.6 Ergodicity and a non-linear Perron-Frobenius theorem

In this section we introduce the notion of ergodicity for upper transition operators and lay bare the link with the Perron-Frobenius theorem. We allow ourselves to be inspired by corresponding notions for non-stationary Markov chains [52, p. 136] and Markov set chains [28] to lead us to the following definition of ergodicity.

Definition 130 (Ergodicity): *An upper transition operator $\overline{\mathrm{T}}$ on $\mathcal{X}$ is called **ergodic** if for all $h \in \mathcal{L}$, the sequence of gambles $\left\{\overline{\mathrm{T}}^k h\right\}_{k \in \mathbb{N}_{\geq 0}}$ converges pointwise to a constant gamble.*

This definition of ergodicity is not exactly the one more commonly encountered in probability or dynamical systems theory, where ergodicity usually refers to the special properties of an invariant measure. Here, ergodicity corresponds to what is usually called “ergodic + aperiodic” in the Markov chain setting.

Consider any gamble $h \in \mathcal{X}$. Ergodicity of an upper transition operator $\overline{\mathrm{T}}$ not only means that the sequence $\overline{\mathrm{T}}^n h$ converges, so the ω-limit set[10] $\omega_{\overline{\mathrm{T}}}(h)$ is a singleton $\{\xi_h\}$, but also that this limit ξ_h is a constant function. Observe that by $\overline{\mathrm{T}}6_{130}$, ξ_h is a fixed point for all $\overline{\mathrm{T}}^k$: $\overline{\mathrm{T}}^k \xi_h = \xi_h$ and therefore $\xi_{\overline{\mathrm{T}}^k h} = \xi_h$ for all $k \in \mathbb{N}_{\geq 0}$. If we denote the constant value of ξ_h by $\overline{\mathrm{P}}_{\overline{\mathrm{T}}}(h)$, then this defines a real functional $\overline{\mathrm{Q}}_{\overline{\mathrm{T}}}$ on $\mathcal{L}(\mathcal{X})$. This functional is an **upper prevision**: it is bounded, sub-additive and non-negatively homogeneous [compare with $\overline{\mathrm{T}}1_{130}$–$\overline{\mathrm{T}}3_{130}$]. It is $\overline{\mathrm{T}}$-invariant in the sense that $\overline{\mathrm{P}}_{\overline{\mathrm{T}}} \circ \overline{\mathrm{T}} = \overline{\mathrm{P}}_{\overline{\mathrm{T}}}$, and it is the only such upper prevision. This shows that our definition of ergodicity is nevertheless in line with the concept more commonly used in systems theory.

Definition 131: *An upper transition operator $\overline{\mathrm{T}}$ on $\mathcal{L}(\mathcal{X})$ is called **Perron Frobenius-like** if there is some real functional $\overline{\mathrm{Q}}_\infty$ on $\mathcal{L}(\mathcal{X})$ such that*

$$\lim_{n\to\infty} \overline{\mathrm{Q}}_1(\overline{\mathrm{T}}^n h) = \overline{\mathrm{Q}}_\infty(h)$$

for all upper previsions $\overline{\mathrm{Q}}_1$ on $\mathcal{L}(\mathcal{X})$ and all $h \in \mathcal{X}$, or in other words, if the sequence of upper previsions $\overline{\mathrm{Q}}_1 \circ \overline{\mathrm{T}}^n$ converges pointwise to some limit that does not depend on the initial value $\overline{\mathrm{Q}}_1$.

As an immediate result, conditions for ergodicity of upper transition operators are conditions for a Perron–Frobenius-like theorem for such transformations to hold.

Theorem 132 (Perron–Frobenius): *An upper transition operator $\overline{\mathrm{T}}$ is Perron-Frobenius-like if and only if it is ergodic, and in that case $\overline{\mathrm{Q}}_\infty = \overline{\mathrm{Q}}_{\overline{\mathrm{T}}}$.*

Proof: Sufficiency. Suppose $\overline{\mathrm{T}}$ is ergodic. Then using the notations established above, $\overline{\mathrm{T}}^n h \to \xi_h$ and therefore $\overline{\mathrm{Q}}(\overline{\mathrm{T}}^n h) \to \overline{\mathrm{Q}}(\xi_h)$ because any upper prevision $\overline{\mathrm{Q}}$ is continuous [compare with $\overline{\mathrm{T}}6_{130}$]. Observe that, since any upper prevision $\overline{\mathrm{Q}}$ is constant-additive [compare with $\overline{\mathrm{T}}4_{130}$ and $\overline{\mathrm{T}}1_{130}$], $\overline{\mathrm{Q}}(\xi_h) = \overline{\mathrm{Q}}_{\overline{\mathrm{T}}}(h)$. Hence $\overline{\mathrm{Q}} \circ \overline{\mathrm{T}}^n \to \overline{\mathrm{Q}}_{\overline{\mathrm{T}}}$, and therefore $\overline{\mathrm{T}}$ is Perron–Frobenius-like, with $\overline{\mathrm{Q}}_\infty = \overline{\mathrm{Q}}_{\overline{\mathrm{T}}}$.

Necessity. Suppose that $\overline{\mathrm{T}}$ is Perron–Frobenius-like, with limit upper prevision $\overline{\mathrm{Q}}_\infty$. Fix any $x \in \mathcal{X}$, and consider the upper prevision $\overline{\mathrm{Q}}_x$ defined by $\overline{\mathrm{Q}}_x(h) := h(x)$ for all $h \in \mathcal{L}(\mathcal{X})$. Then by assumption $\overline{\mathrm{T}}^n h(x) = \overline{\mathrm{Q}}_x(\overline{\mathrm{T}}^n h) \to \overline{\mathrm{Q}}_\infty(h)$. Since this holds for all $x \in \mathcal{X}$, we see that $\overline{\mathrm{T}}$ is ergodic with $\overline{\mathrm{Q}}_{\overline{\mathrm{T}}} = \overline{\mathrm{Q}}_\infty$. □

[10] Here, $\omega_{\overline{\mathrm{T}}}(h)$ stands for the ω-limit set of h, which is the set of cluster points of the orbit $\{\overline{\mathrm{T}}^n h\}_{n\in\mathbb{N}_{\geq 0}}$. In other words, $g \in \omega_{\overline{\mathrm{T}}}(h)$ if and only if there exists a strictly increasing sequence of natural numbers $\{n_k\}_{k\in\mathbb{N}_{\geq 0}}$ such that $g = \lim_{k\to\infty} \overline{\mathrm{T}}^{n_k} h$.

It follows from the discussion in Section 3.2[129] that $\bigcup_{h\in\mathcal{L}(\mathcal{X})}\omega_{\overline{\mathrm{T}}}(h)$ is the set of all periodic points of $\overline{\mathrm{T}}$—**a periodic point** being an element $h\in\mathcal{L}(\mathcal{X})$ for which there is some $n\in\mathbb{N}_{\geq0}$ for which $\overline{\mathrm{T}}^n h = h$. Because of $\overline{\mathrm{T}}4$[130], this set contains all constant maps. We now see that for $\overline{\mathrm{T}}$ to be ergodic, this set cannot contain any other maps.

Proposition 133: *An upper transition operator $\overline{\mathrm{T}}$ is ergodic if and only if all of its periodic points are constant maps.*

3.7 Characterisation of ergodicity

We now turn to the issue of determining in practice whether an upper transition operator is ergodic. In the case of finite-state, discrete-time Markov chains, a nice approach to deciding upon ergodicity, based on the the accessibility relation, was given by Kemeny and Snell [35, § 1.4].

In the case of finite-state, discrete-time precise Markov chains, the accessibility of states, gives us clues about the ergodicity of the Markov chain. For such a Markov chain to be ergodic, it is necessary and sufficient that it is **top class regular** [14], meaning that: (i) there is only one **maximal** or **undominated** communication class—elements of a maximal communication class have no access to states not in that class—, in which case we call this unique maximal class $\mathcal{R}$ the **top class**; and (ii) the top class $\mathcal{R}$ should be regular, meaning that after some time k, all elements of this class become accessible to each other in any number of steps: for all x and y in $\mathcal{R}$ and for all $n\geq k$, $x\overset{n}{\to}y$.

For upper transition operators, it turns out that top class regularity is a necessary condition for ergodicity. However, top class regularity is by itself not a sufficient condition: we need some guarantee that the top class will eventually be reached—a requirement that is automatically fulfilled in finite-state discrete-time precise Markov chains.

Proposition 134: *An upper transition operator $\overline{\mathrm{T}}$ is ergodic if and only if it is **regularly absorbing**, meaning that it satisfies the following properties:*

(TCR) *it is **top class regular**:*

$$\mathcal{R} := \left\{x\in\mathcal{X}\colon (\exists n\in\mathbb{N}_{\geq0})(\forall k\geq n)\min\overline{\mathrm{T}}^k I_{\{x\}}>0\right\}\neq\varnothing,$$

(TCA) *it is **top class absorbing**: with* $\mathcal{R}^c := \mathcal{X} \setminus \mathcal{R}$,

$$(\forall y \in \mathcal{R}^c)(\exists n \in \mathbb{N}_{\geq 0})\,\overline{\mathrm{T}}^n I_{\mathcal{R}^c}(y) < 1.$$

For a proof that (TCR) is equivalent to $\mathcal{R} \neq \emptyset$, we refer to [14, Prop. 4.3]. (TCA) means that for every element y not in the top class, there is some finite number of steps n after which the top class can be reached with a strictly positive **lower** probability $1 - \overline{\mathrm{T}}^n I_{\mathcal{R}^c}(y)$.

Proof: (TCR) $\wedge$ (TCA) $\Rightarrow$ (ER). Consider any fixed point ξ of $\overline{\mathrm{T}}^k$, where $k \in \mathbb{N}_{\geq 0}$ and observe, by $\overline{\mathrm{T}}5_{130}$ and $\overline{\mathrm{T}}4_{130}$, that $\min \xi \leq \min \overline{\mathrm{T}}\xi \leq \min \overline{\mathrm{T}}^2\xi \leq \ldots \leq \min \overline{\mathrm{T}}^k \xi = \min \xi$ whence for any $p \in \mathbb{N}_{\geq 0}$,

$$\min \xi = \min \overline{\mathrm{T}}^p \xi \quad \text{and similarly} \quad \max \xi = \max \overline{\mathrm{T}}^p \xi. \tag{3.32}$$

We infer from Proposition 133 that we have to show that ξ is constant. Using $\overline{\mathrm{T}}5_{130}$, $\overline{\mathrm{T}}4_{130}$, $\overline{\mathrm{T}}3_{130}$ and Eq. 3.32 we construct from $\overline{\mathrm{T}}^p \xi \geq \min \overline{\mathrm{T}}^p \xi + [\overline{\mathrm{T}}^p \xi(x) - \min \overline{\mathrm{T}}^p \xi] I_{\{x\}} = \min \xi + [\overline{\mathrm{T}}^p \xi(x) - \min \xi] I_{\{x\}}$ the following inequality, which holds for all $n, p \in \mathbb{N}_{\geq 0}$ and all $x \in \mathcal{X}$:

$$\overline{\mathrm{T}}^{n+p} \xi \geq \min \xi + \left[\overline{\mathrm{T}}^p \xi(x) - \min \xi\right] \overline{\mathrm{T}}^n I_{\{x\}}.$$

Hence, by taking the minimum on both sides of this inequality and using Equation (3.32), we find that

$$0 \geq \left[\overline{\mathrm{T}}^p \xi(x) - \min \xi\right] \min \overline{\mathrm{T}}^n I_{\{x\}}.$$

We infer from (TCR) that by taking n large enough, we can ensure that $\min \overline{\mathrm{T}}^n I_{\{x\}} > 0$ whence for any $p \in \mathbb{N}_{\geq 0}$ and $x \in \mathcal{R}$

$$0 = \left[\overline{\mathrm{T}}^p \xi(x) - \min \xi\right],$$

so we already find that $\overline{\mathrm{T}}^p \xi(x) = \min \xi$ for all $p \in \mathbb{N}_{\geq 0}$ and $x \in \mathcal{R}$.

If there is some $p \in \mathbb{N}_{\geq 0}$ such that $\overline{\mathrm{T}}^p \xi$ reaches its maximum on $\mathcal{R}$, then we infer from Eq. (3.32) that $\max \overline{\mathrm{T}}^p \xi = \max \xi$ which has to be equal to $\min \xi$ to satisfy the inequality, so ξ is indeed constant. Let us therefore assume that the maximum of $\overline{\mathrm{T}}^p \xi$ is not reached in $\mathcal{R}$. Using $\overline{\mathrm{T}}5_{130}$, $\overline{\mathrm{T}}4_{130}$, $\overline{\mathrm{T}}3_{130}$ and Equation (3.32), we construct from $\xi \leq \max \xi - [\max \xi - \max_{x \in \mathcal{R}} \xi(x)] I_{\mathcal{R}}$ and $-I_{\mathcal{R}} = I_{\mathcal{R}^c} - 1$ the following inequality, which holds for all $n \in \mathbb{N}_{\geq 0}$:

$$\overline{\mathrm{T}}^n \xi \leq \max \xi + \left[\max \xi - \max_{x \in \mathcal{R}} \xi(x)\right] (\overline{\mathrm{T}}^n I_{\mathcal{R}^c} - 1).$$

By taking the maximum over $\mathcal{R}^c$ on both sides of this inequality and under the made assumption that the maximum is never reached on $\mathcal{R}$, we get

$$0 = \max_{y\in\mathcal{R}^c} \overline{\mathrm{T}}^n \xi(y) - \max \xi \leq \left[\max \xi - \max_{x\in\mathcal{R}} \xi(x)\right]\left(\max_{y\in\mathcal{R}^c} \overline{\mathrm{T}}^n I_{\mathcal{R}^c}(y) - 1\right).$$

For each $y \in \mathcal{R}^c$, consider some $n_y \in \mathbb{N}_{\geq 0}$ such that $\overline{\mathrm{T}}^{n_y} I_{\mathcal{R}^c}(y) < 1$, and let $n := \max_{y\in\mathcal{R}^c} n_y$. Then we see that for every $y \in \mathcal{R}^c$:

$$\begin{aligned}
\overline{\mathrm{T}}^n I_{\mathcal{R}^c}(y) &= \overline{\mathrm{T}}^{n_y}[(I_{\mathcal{R}} + I_{\mathcal{R}^c})\overline{\mathrm{T}}^{n-n_y} I_{\mathcal{R}^c}](y) \\
&= \overline{\mathrm{T}}^{n_y}[I_{\mathcal{R}^c}\overline{\mathrm{T}}^{n-n_y} I_{\mathcal{R}^c}](y) \\
&\leq \overline{\mathrm{T}}^{n_y} I_{\mathcal{R}^c}(y) < 1.
\end{aligned}$$

The second equality follows from the fact that $I_{\mathcal{R}}\overline{\mathrm{T}}^{n-n_y} I_{\mathcal{R}^c} = 0$: an element in the top class $\mathcal{R}$ has no access to any element outside of it; and the first inequality follows from $I_{\mathcal{R}^c} \leq 1$ and $\overline{\mathrm{T}}5_{130}$. But this means that $\max_{y\in\mathcal{R}^c} \overline{\mathrm{T}}^n I_{\mathcal{R}^c}(y) - 1 < 0$ and consequently $\max \xi = \max_{x\in\mathcal{R}} \xi(x) = \min \xi$.

(ER) $\Rightarrow$ (TCR) $\wedge$ (TCA). We will use contraposition and show first that $\neg$(TCR) $\Rightarrow$ $\neg$(ER). Then we will show that $\neg$(TCA) $\wedge$ (TCR) $\Rightarrow$ $\neg$(ER).

$\neg$(TCR) $\Rightarrow$ $\neg$(ER). Not being top class regular means that $\mathcal{R} = \emptyset$, which is equivalent to

$$(\forall x \in \mathcal{X})(\forall n \in \mathbb{N}_{\geq 0})(\exists k \geq n)(\exists z \in \mathcal{X})\overline{\mathrm{T}}^k I_{\{x\}}(z) = 0.$$

Since we infer from $I_{\{x\}} \geq 0$ and $\overline{\mathrm{T}}1_{130}$ that $\overline{\mathrm{T}}^k I_{\{x\}} \geq 0$, this leads us to conclude that $\liminf_{n\to\infty} \min \overline{\mathrm{T}}^n I_{\{x\}} = 0$. But for any $n \in \mathbb{N}_{\geq 0}$, $\overline{\mathrm{T}}^{n+1} I_{\{x\}} = \overline{\mathrm{T}}(\overline{\mathrm{T}}^n I_{\{x\}}) \geq \min \overline{\mathrm{T}}^n I_{\{x\}}$ by $\overline{\mathrm{T}}1_{130}$, and therefore also $\min \overline{\mathrm{T}}^{n+1} I_{\{x\}} \geq \min \overline{\mathrm{T}}^n I_{\{x\}}$. This implies that the sequence $\min \overline{\mathrm{T}}^n I_{\{x\}}$ is non-decreasing, and bounded above [by 1], and therefore convergent. This shows that

$$(\forall x \in \mathcal{X}) \lim_{n\to\infty} \min \overline{\mathrm{T}}^n I_{\{x\}} = 0. \tag{3.33}$$

We also infer from $\overline{\mathrm{T}}1_{130}$ and $\overline{\mathrm{T}}2_{130}$ that $1 = \overline{\mathrm{T}}^k I_{\mathcal{X}} \leq \sum_{x\in\mathcal{X}} \overline{\mathrm{T}}^k I_{\{x\}}$. Since the cardinality $|\mathcal{X}|$ of the state space is finite, this means that for all $z \in \mathcal{X}$ and all $n \in \mathbb{N}_{\geq 0}$ there is some $x \in \mathcal{X}$ such that $\overline{\mathrm{T}}^n I_{\{x\}}(z) \geq 1/|\mathcal{X}|$. This tells us that $\max \overline{\mathrm{T}}^n I_{\{x\}} \geq 1/|\mathcal{X}|$. Since we can infer from a similar argument as before that the sequence $\max \overline{\mathrm{T}}^n I_{\{x\}}$ converges, this tells us that

$$(\forall x \in \mathcal{X}) \lim_{n\to\infty} \max \overline{\mathrm{T}}^n I_{\{x\}} \geq \frac{1}{|\mathcal{X}|}. \tag{3.34}$$

Combining Equations (3.33) and (3.34) tells us that

$$\lim_{n\to\infty} \left(\max \overline{\mathrm{T}}^n I_{\{x\}} - \min \overline{\mathrm{T}}^n I_{\{x\}}\right) > 0,$$

so $\overline{\mathrm{T}}$ cannot be ergodic.

$\neg$(TCA) $\wedge$ (TCR) $\Rightarrow$ $\neg$(ER). Since $\overline{\mathrm{T}}$ is not top class absorbing, we know that there is some $y \in \mathcal{R}^c$ such that $\overline{\mathrm{T}}^n I_{\mathcal{R}^c}(y) = 1$ for all $n \in \mathbb{N}_{\geq 0}$. As the top class $\mathcal{R}$ is non-empty, we know that there is some $x \in \mathcal{R}$, and this x has no access to any state outside the maximal communication class $\mathcal{R}$: $\overline{\mathrm{T}}^n I_{\mathcal{R}^c}(x) = 0$ for all $n \in \mathbb{N}_{\geq 0}$. Consequently

$$\lim_{n\to\infty} \left(\max \overline{\mathrm{T}}^n I_{\mathcal{R}^c} - \min \overline{\mathrm{T}}^n I_{\mathcal{R}^c}\right) = 1 - 0 > 0,$$

so $\overline{\mathrm{T}}$ cannot be ergodic. □

3.8 Ergodicity checking in practice

3.8.1 Checking for top class regularity

Checking for top class regularity directly using the definition would involve calculating for every state x the maps $\overline{\mathrm{T}} I_{\{x\}}, \overline{\mathrm{T}}^2 I_{\{x\}}, \ldots, \overline{\mathrm{T}}^n I_{\{x\}}$ until a first number $n = n_x$ is found for which $\min \overline{\mathrm{T}}^{n_x} I_{\{x\}} > 0$. Unfortunately, it is not clear whether this procedure is guaranteed to terminate, or whether we can stop checking after a fixed number of iterations. However, it is clear from Proposition 120138 that the accessibility relation $\dot{\rightarrow}$ of the accessibility graph $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ corresponding to $\overline{\mathrm{T}}$, is exactly the accessibility relation $\rightarrow$ belonging to the upper transition operator $\overline{\mathrm{T}}$. This means that checking for the existence of a single top class of $\overline{\mathrm{T}}$, corresponds to asserting whether there is only one final class $\mathcal{R}$ in $\overline{\mathcal{G}}(\overline{\mathrm{T}})$. Once we have found the top class $\mathcal{R}$, we focus on the subgraph $\overline{\mathcal{G}}(\overline{\mathrm{T}})\big|_{\mathcal{R}}$ which is the upper accessibility graph $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ restricted to $\mathcal{R}$. Property 128141 tells us that checking for regularity of the top class means that we have to check whether the cyclicity of $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ is equal to 1.

The relation between $\overline{\mathrm{T}}$ and its graph $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ is a purely qualitative one: the exact quantitative value of the upper transition probabilities between two states x and y is not important at all. What is important is whether there is a possible transition between two states. This means that appropriately replacing the upper transition operator $\overline{\mathrm{T}}$ with a classical, linear transition operator, or its associated transition matrix M, will still lead to the same results.

Definition 135: *A stochastic matrix $M \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$* ***represents*** *an upper transition operator $\overline{\mathrm{T}}$ on $\mathcal{X}$ if $M_{x,y} > 0 \Leftrightarrow \overline{\mathrm{T}} I_{\{y\}}(x) > 0$ for all x and y in $\mathcal{X}$.*

It is clear that any stochastic matrix M that represents $\overline{\mathrm{T}}$ will result into the same graph $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ and will therefore lead to the same conclusions with respect to top class regularity. For stochastic matrices however, a final class corresponds to an irreducible stochastic submatrix, and aperiodicity corresponds to the absence of eigenvalues with modulus 1 apart from 1 (with multiplicity 1) itself.

Proposition 136 (Top class regularity): *Consider an upper transition operator* $\overline{\mathrm{T}}$*. Then the following statements are equivalent:* (i) $\overline{\mathrm{T}}$ *is top class regular;* (ii) M *represents* $\overline{\mathrm{T}}$ *and is regular;* (iii) M *represents* $\overline{\mathrm{T}}$ *and* M *has exactly one eigenvalue with modulus* 1*; and* (iv) $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ *has exactly one final class* $\mathcal{R}$ *and* $\overline{\mathcal{G}}(\overline{\mathrm{T}})\big|_{\mathcal{R}}$ *has cyclicity* 1*.*

▷ **Example 137:** Let $\mathcal{X} := \{x, y\}$ and $\overline{\mathrm{T}}f := f(x)I_{\{x\}} + \max\{f(x), f(y)\}I_{\{y\}}$ for all $f \in \mathcal{X}$. Then $\overline{\mathrm{T}}I_{\{x\}} = I_{\mathcal{X}}$ whence $x \overset{1}{\to} x$ and $y \overset{1}{\to} x$ and $\overline{\mathrm{T}}I_{\{y\}} = I_{\{y\}}$ whence $y \overset{1}{\to} y$. The graph $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ is then given by

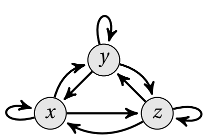
Clearly $\{x\}$ is the unique final strongly connected component of $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ and as it is a singleton, it has cyclicity one. We conclude that $\overline{\mathrm{T}}$ is top class regular.

In the next example we focus on a simple upper transition operator that is not piecewise affine. It does not therefore fall within the scope of Akian and Gaubert's algorithm, which we will discuss in Section 3.10$_{154}$.

▷ **Example 138:** Consider the map

$$\overline{\mathrm{T}}\colon \mathbb{R}^3 \to \mathbb{R}^3 \colon f \to \overline{f} + \frac{\|f - \overline{f}\|_2}{\sqrt{3}} \begin{pmatrix} \alpha_x \\ \alpha_y \\ \alpha_z \end{pmatrix}$$

where $\overline{f} := (f_x + f_y + f_z)/3$ for $f = \begin{pmatrix} f_x & f_y & f_z \end{pmatrix}^T$ and the parameters α_x, α_y and α_z are any real numbers in $[0, 1/\sqrt{2}]$. It is not difficult to check that this $\overline{\mathrm{T}}$ is indeed an upper transition operator, but it is obviously not piecewise affine. Independently of the value of α_x, α_y and α_z, the upper accessibility graph of this map is given by:



The entire graph is strongly connected, and it has cyclicity one. This implies that $\overline{\mathrm{T}}$ is not only top class regular, but also ergodic, according to Proposition 134$_{145}$.

3.8.2 Checking for top class absorption

We now present a computationally efficient procedure to check for top class absorption.

Proposition 139 (Top class absorption): *Let $\overline{\mathrm{T}}$ be an upper transition operator with regular top class $\mathcal{R}$. Consider the nested sequence of subsets of $\mathcal{R}^c$ defined by the iterative scheme:*

$$\begin{aligned} A_0 &:= \mathcal{R}^c \\ A_{n+1} &:= \left\{a \in A_n : \overline{\mathrm{T}} I_{A_n}(a) = 1\right\}, \quad n \geq 0. \end{aligned}$$

After $k \leq |\mathcal{R}^c|$ iterations, we reach $A_k = A_{k+1}$. Then $\overline{\mathrm{T}}$ is top class absorbing if and only if $A_k = \emptyset$.

Proof: We start by showing inductively that under the given assumptions, the statement

$$\begin{aligned} H_n: \quad & \overbrace{I_{A_n} \overline{\mathrm{T}}^n I_{\mathcal{R}^c} = I_{A_n}}^{(\alpha)} \text{ and} \\ & \underbrace{(\forall a \in A_{n+1}^c) \overline{\mathrm{T}} I_{A_n}(a) < 1}_{(\beta)} \text{ and } \underbrace{(\forall a \in A_n^c) \overline{\mathrm{T}}^n I_{\mathcal{R}^c}(a) < 1}_{(\gamma)} \end{aligned}$$

holds for all $n \geq 0$. We first prove that the statement H_n holds for $n = 0$. The first and third statements of H_0 hold trivially. For the second statement, we have to prove that $\overline{\mathrm{T}} I_{A_0}(a) < 1$ for all $a \in A_1^c = A_0^c \cup (A_0 \setminus A_1)$. On $A_0 \setminus A_1$, the desired inequality holds by definition. On $A_0^c = \mathcal{R}$ it holds because there $\overline{\mathrm{T}} I_{A_0}$ is zero: no state in the top class $\mathcal{R}$ has access to any state outside it.

Next, we prove that $H_n \Rightarrow H_{n+1}$. First of all (α),

$$\overline{\mathrm{T}}^{n+1} I_{A_0} = \overline{\mathrm{T}}(\overline{\mathrm{T}}^n I_{A_0}) = \overline{\mathrm{T}}[I_{A_n} \overline{\mathrm{T}}^n I_{A_0} + I_{A_n^c} \overline{\mathrm{T}}^n I_{A_0}] = \overline{\mathrm{T}}[I_{A_n} + I_{A_n^c} \overline{\mathrm{T}}^n I_{A_0}], \tag{3.35}$$

where the last equality follows from the induction hypothesis H_n. It follows from the definition of A_{n+1} that $I_{A_{n+1}} \overline{\mathrm{T}} I_{A_n} = I_{A_{n+1}}$, and therefore

$$\begin{aligned} I_{A_{n+1}} &= I_{A_{n+1}} \overline{\mathrm{T}}[I_{A_n} + I_{A_n^c} \overline{\mathrm{T}}^n I_{A_0} - I_{A_n^c} \overline{\mathrm{T}}^n I_{A_0}] \\ &\leq I_{A_{n+1}} \overline{\mathrm{T}}[I_{A_n} + I_{A_n^c} \overline{\mathrm{T}}^n I_{A_0}] + I_{A_{n+1}} \overline{\mathrm{T}}[-I_{A_n^c} \overline{\mathrm{T}}^n I_{A_0}] \\ &= I_{A_{n+1}} \overline{\mathrm{T}}^{n+1} I_{A_0} + I_{A_{n+1}} \overline{\mathrm{T}}[-I_{A_n^c} \overline{\mathrm{T}}^n I_{A_0}] \\ &\leq I_{A_{n+1}} \overline{\mathrm{T}}^{n+1} I_{A_0} \leq I_{A_{n+1}}, \end{aligned}$$

where the first inequality follows from $\overline{\mathrm{T}}2_{130}$, the second inequality follows from the fact that $-I_{A_n^c} \overline{\mathrm{T}}^n I_{A_0} \leq 0$ and therefore $I_{A_{n+1}} \overline{\mathrm{T}}[-I_{A_n^c} \overline{\mathrm{T}}^n I_{A_0}] \leq 0$ [use $\overline{\mathrm{T}}1_{130}$ and $\overline{\mathrm{T}}5_{130}$], and the third from $\overline{\mathrm{T}}^{n+1} I_{A_0} \leq 1$ [use $\overline{\mathrm{T}}5_{130}$]. The second equality follows from Equation (3.35). Hence indeed $I_{A_{n+1}} = I_{A_{n+1}} \overline{\mathrm{T}}^{n+1} I_{A_0}$.

(β) Next, observe that $A^c_{n+2} = A^c_{n+1} \cup (A_{n+1} \setminus A_{n+2})$. By definition, $\overline{\mathrm{T}} I_{A_{n+1}}(a) < 1$ for all $a \in A_{n+1} \setminus A_{n+2}$. It also follows from the induction hypothesis H_n that $\overline{\mathrm{T}} I_{A_n}(a) < 1$ for all $a \in A^c_{n+1}$. But since $A_{n+1} \subseteq A_n$, it follows from $\overline{\mathrm{T}}5_{130}$ that $\overline{\mathrm{T}} I_{A_{n+1}} \leq \overline{\mathrm{T}} I_{A_n}$, and therefore also $\overline{\mathrm{T}} I_{A_{n+1}}(a) < 1$ for all $a \in A^c_{n+1}$. Hence indeed $\overline{\mathrm{T}} I_{A_{n+1}}(a) < 1$ for all $a \in A^c_{n+2}$.

(γ) To finish the induction proof, let $\beta := \max_{a \in A^c_n} \overline{\mathrm{T}}^n I_{\mathcal{R}^c}(a)$, then $\beta < 1$ by the induction hypothesis H_n. We then infer from Equation (3.35) that

$$\overline{\mathrm{T}}^{n+1} I_{A_0} = \overline{\mathrm{T}}[I_{A_n} + I_{A^c_n} \overline{\mathrm{T}}^n I_{A_0}] \leq \overline{\mathrm{T}}[I_{A_n} + \beta I_{A^c_n}] = \overline{\mathrm{T}}[\beta + (1-\beta) I_{A_n}] = \beta + (1-\beta)\overline{\mathrm{T}} I_{A_n}.$$

Consider any $a \in A^c_{n+1}$, then $\overline{\mathrm{T}} I_{A_n}(a) < 1$ by the induction hypothesis H_n, and therefore $\overline{\mathrm{T}}^{n+1} I_{A_0}(a) \leq \beta + (1-\beta)\overline{\mathrm{T}} I_{A_n}(a) < 1$ since also $\beta < 1$. We conclude that H_{n+1} holds too.

To continue the proof, we observe that $A_0, A_1, \ldots, A_n, \ldots$ is a non-increasing sequence, and that A_0 is finite. This implies that there must be some first $k \in \mathbb{N}_{\geq 0}$ such that $A_{k+1} = A_k$. Clearly, $k \leq |A_0|$. We now prove by induction that $G_n : I_{A_k} \overline{\mathrm{T}}^{n+k} I_{A_0} = I_{A_k}$ for all $n \geq 0$. The statement G_n clearly holds for $n = 0$: it follows directly from H_k. We show that $G_n \Rightarrow G_{n+1}$. First of all,

$$\overline{\mathrm{T}}^{n+k+1} I_{A_0} = \overline{\mathrm{T}}(\overline{\mathrm{T}}^{n+k} I_{A_0}) = \overline{\mathrm{T}}[I_{A_k} \overline{\mathrm{T}}^{n+k} I_{A_0} + I_{A^c_k} \overline{\mathrm{T}}^{n+k} I_{A_0}] = \overline{\mathrm{T}}[I_{A_k} + I_{A^c_k} \overline{\mathrm{T}}^{n+k} I_{A_0}],$$

where the last equality follows from the induction hypothesis G_n. As before, it follows from the definition of A_{k+1} that $I_{A_{k+1}} \overline{\mathrm{T}} I_{A_k} = I_{A_{k+1}}$, and therefore $I_{A_k} \overline{\mathrm{T}} I_{A_k} = I_{A_k}$ (remember that $A_{k+1} = A_k$), so

$$\begin{aligned} I_{A_k} &= I_{A_k} \overline{\mathrm{T}}[I_{A_k} + I_{A^c_k} \overline{\mathrm{T}}^{n+k} I_{A_0} - I_{A^c_k} \overline{\mathrm{T}}^{n+k} I_{A_0}] \\ &\leq I_{A_k} \overline{\mathrm{T}}[I_{A_n k} + I_{A^c_k} \overline{\mathrm{T}}^{n+k} I_{A_0}] + I_{A_k} \overline{\mathrm{T}}[-I_{A^c_k} \overline{\mathrm{T}}^{n+k} I_{A_0}] \\ &= I_{A_k} \overline{\mathrm{T}}^{n+k+1} I_{A_0} + I_{A_k} \overline{\mathrm{T}}[-I_{A^c_k} \overline{\mathrm{T}}^{n+k} I_{A_0}] \leq I_{A_k} \overline{\mathrm{T}}^{n+k+1} I_{A_0} \leq I_{A_k}, \end{aligned}$$

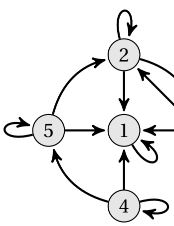
where the first inequality follows from $\overline{\mathrm{T}}2_{130}$ and the second inequality follows from the fact that $-I_{A^c_k} \overline{\mathrm{T}}^{n+k} I_{A_0} \leq 0$ and therefore $I_{A_k} \overline{\mathrm{T}}[-I_{A^c_k} \overline{\mathrm{T}}^{n+k} I_{A_0}] \leq 0$ [use $\overline{\mathrm{T}}1_{130}$ and $\overline{\mathrm{T}}5_{130}$], and the third from $\overline{\mathrm{T}}^{n+k+1} I_{A_0} \leq 1$ [use $\overline{\mathrm{T}}5_{130}$]. Hence indeed $I_{A_k} = I_{A_k} \overline{\mathrm{T}}^{n+k+1} I_{A_0}$.

There are now two possibilities. The first is that $A_k \neq \emptyset$. It follows from the arguments above that for any element a of A_k, $\overline{\mathrm{T}}^\ell I_{\mathcal{R}^c}(a) = 1$ for all $\ell \in \mathbb{N}_{\geq 0}$, which implies that $\overline{\mathrm{T}}$ cannot be top class absorbing. The second possibility is that $A_k = \emptyset$. It follows from the argument above that $\overline{\mathrm{T}}^k I_{\mathcal{R}^c}(a) < 1$ for all $a \in A^c_k = \mathcal{X}$ which implies that $\overline{\mathrm{T}}$ is top class absorbing. □

▷ **Example 140:** Define $\overline{\mathrm{T}}f = \max\{Mf : L \le M \le U \text{ and } M \text{ stochastic}\}$ where L and U are given by

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1/4 & 1/4 & 0 & 0 \\ 1/2 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/4 \end{pmatrix} \text{ and } U = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 3/4 & 1/2 & 0 & 0 \\ 3/4 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1/4 & 3/4 & 0 & 0 & 1/4 \end{pmatrix}.$$

The corresponding upper accessibility graph $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ is given by



where $\{1\}$ is corresponds to the unique strongly connected component that is final. As it is a singleton, it has cyclicity one, so there is a regular top class $\mathcal{R} = \{1\}$.

To check for top class absorption, we start iterating:

step 1: $\overline{\mathrm{T}}I_{\mathcal{R}^c} = \begin{pmatrix} 0 & 1 & 1/2 & 1 & 1 \end{pmatrix}^T$ whence $I_{A_1} = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \end{pmatrix}^T$,

step 2: $\overline{\mathrm{T}}I_{A_1} = \begin{pmatrix} 0 & 3/4 & 1/2 & 1 & 1 \end{pmatrix}^T$ whence $I_{A_2} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \end{pmatrix}^T$,

step 3: $\overline{\mathrm{T}}I_{A_2} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1/4 \end{pmatrix}^T$ whence $I_{A_3} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \end{pmatrix}^T$,

step 4: $\overline{\mathrm{T}}I_{A_3} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \end{pmatrix}^T$ whence $I_{A_4} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \end{pmatrix}^T$.

Because $A_4 = A_3 \neq \emptyset$ we conclude that $\overline{\mathrm{T}}$ is not top class absorbing and therefore not ergodic.

3.9 Coefficient of ergodicity

It is clear that ergodicity would follow immediately from Banach's fixed point theorem if $\overline{\mathrm{T}}$ were contractive instead of non-expansive. With this in mind, one might think that conditions for ergodicity might coincide with contractiveness of $\overline{\mathrm{T}}$. This is not true. Take, for example, the particular upper transition operator $\overline{\mathrm{T}} = I_{\mathcal{X}} \max$, which is not contractive, but, by Proposition 133_{145}, clearly ergodic.

In addition to requiring the sequence $\{\overline{\mathrm{T}}^k h\}_{k \in \mathbb{N}_{\geq 0}}$ to converge pointwise, ergodicity also requires that the sequence converges to a constant gamble.

Therefore, $\max \overline{\mathrm{T}}^n f - \min \overline{\mathrm{T}}^n f \to 0$ when $n \to \infty$. It seems therefore to be more natural to focus on the so-called **variation pseudo-norm** defined by:

$$\|f\|_{\mathrm{v}} := \max f - \min f.$$

Under this pseudo-norm, upper transition operators will again be non-expansive. The extra condition that makes the map $\overline{\mathrm{T}}$ contractive is expressed by Škulj and Hable [61] in terms of the coefficient of ergodicity. It is a standard trick, see Nussbaum's monograph [44], to use Hilbert's projective metric to show contraction. The variation norm we define can now be seen as an additive version of Hilbert's projective metric.

Proposition 141: *If we define the coefficient of ergodicity of an upper transition operator* $\overline{\mathrm{T}}$ *as*

$$\rho(\overline{\mathrm{T}}) := \max\left\{ \left\|\overline{\mathrm{T}}h\right\|_{\mathrm{v}} : 0 \leq h \leq 1 \right\}, \tag{3.36}$$

then $\overline{\mathrm{T}}$ *is ergodic if* $\rho(\overline{\mathrm{T}}^m) < 1$ *for some* $m \in \mathbb{N}_{\geq 0}$.

Proof: Consider any $f \in \mathcal{X}$. It follows by repeatedly applying $\overline{\mathrm{T}}5_{130}$, $\overline{\mathrm{T}}3_{130}$ and $\overline{\mathrm{T}}4_{130}$ that for all $k \in \mathbb{N}_{\geq 0}$:

$$\min f \leq \min \overline{\mathrm{T}}^k f \leq \min \overline{\mathrm{T}}^{k+1} f \leq \max \overline{\mathrm{T}}^{k+1} f \leq \max \overline{\mathrm{T}}^k f \leq \max f. \tag{3.37}$$

This tells us that the sequence $\min \overline{\mathrm{T}}^k f$ is non-decreasing and bounded above. It therefore converges to some real number m. Similarly, the sequence $\max \overline{\mathrm{T}}^k f$ is non-increasing and bounded below, and therefore converges to some real number M. It is also clear from Equation (3.37) that $m \leq M$. Suppose that there is some $p \in \mathbb{N}_{\geq 0}$ such that $\rho(\overline{\mathrm{T}}^p) < 1$. Then we have to prove that $m = M$, which is what we now set out to do.

Since $0 \leq (f - \min f)/\|f\|_{\mathrm{v}} \leq 1$, we infer from Equation (3.36), ($\overline{\mathrm{T}}3$) and ($\overline{\mathrm{T}}4$) that

$$\frac{\left\|\overline{\mathrm{T}}f\right\|_{\mathrm{v}}}{\|f\|_{\mathrm{v}}} = \left\|\overline{\mathrm{T}}\frac{f - \min f}{\|f\|_{\mathrm{v}}}\right\|_{\mathrm{v}} \leq \rho(\overline{\mathrm{T}}),$$

and therefore also

$$\left\|\overline{\mathrm{T}}^k f\right\|_{\mathrm{v}} \leq \rho(\overline{\mathrm{T}}^k)\|f\|_{\mathrm{v}} \text{ for all } k \in \mathbb{N}_{\geq 0}. \tag{3.38}$$

Then applying Equation (3.38) repeatedly tells us that for the upper transition operator $\Lambda := \overline{\mathrm{T}}^p$:

$$\left\|\Lambda^n f\right\|_{\mathrm{v}} \leq \rho(\overline{\mathrm{T}}^p)^n \|f\|_{\mathrm{v}} \text{ for all } n \in \mathbb{N}_{\geq 0}.$$

But this implies that $\max \Lambda^n f - \min \Lambda^n f = \|\Lambda^n f\|_{\mathrm{v}} \to 0$. Since we know from the arguments above that $\max \Lambda^n f \to M$ and $\min \Lambda^n f \to m$, this implies that indeed $m = M$. □

Not only does the coefficient of ergodicity allow us to decide in favour of ergodicity, but by Equation (3.38) it also gives a numerical bound on the speed of convergence. The main problem however is that, in the worst case, in order to check for ergodicity in this manner, we need to calculate the coefficient of ergodicity of $\overline{\mathrm{T}}^k$ for powers k up to $g(|\mathcal{X}|)$, where g is Landau's function. This renders this approach, as described in Section 3.8$_{148}$, impractical from a computational point of view, making our approach preferable.

The following section describes yet another alternate approach for checking ergodicity.

3.10 The critical graph versus the upper accessibility graph

The aim of Akian and Gaubert's paper [1] is to determine, for convex, monotone and non-expansive maps Φ, combinatorial bounds on orbit lengths of the described maps. Although the scope of Akian and Gaubert's paper is different, it overlaps to some extent with our work on the limit behaviour of upper transition operators. Akian and Gaubert try to describe the entire (additive) eigenspace of the map Φ. Their tool of choice for doing that is what they call the **critical graph** $\mathcal{G}^c(\Phi)$ of the map Φ. It is defined as the final graph $\mathcal{G}^f(\partial\Phi(v))$ of the subdifferential $\partial\Phi$ of Φ evaluated in an (additive) eigenvector v. Akian and Gaubert define the subdifferential of the operator Φ evaluated in any vector v as

$$\partial\Phi(v) := \left\{M \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|} : (\forall f \in \mathbb{R}^{|\mathcal{X}|})\Phi f - \Phi v \geq M(f - v)\right\}.$$

They show that the matrices M that belong to $\partial\Phi(v)$ are necessarily stochastic matrices.

Let us now consider what happens in the special case that Φ is an upper transition operator $\overline{\mathrm{T}}$, in order to better understand the relationship between their approach and ours. Given the constant additivity of $\overline{\mathrm{T}}$ we can choose any constant gamble as an (additive) eigenvector to calculate the critical graph. To make things as simple as possible, we opt for the zero gamble. The subdifferential of $\overline{\mathrm{T}}$ evaluated in this additive eigen vector then becomes

$$\partial\overline{\mathrm{T}}(0)=\left\{M\in\mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}:(\forall f\in\mathbb{R}^{|\mathcal{X}|})\overline{\mathrm{T}}f\geq M(f)\right\}=\mathcal{T},$$

which is the closed convex set of transition matrices that corresponds with the upper transition operator $\overline{\mathrm{T}}$, as defined by Equation (3.19). The critical graph $\mathcal{G}^c(\overline{\mathrm{T}})=\mathcal{G}^f(\partial\overline{\mathrm{T}}(0))=\mathcal{G}^f(\mathcal{T})$ is then (defined as) the union of all the final graphs of the stochastic matrices belonging to $\mathcal{T}$. A final graph of a stochastic matrix can be found by interpreting this stochastic matrix as an adjacency matrix and restricting the corresponding graph to its final classes (see also the discussion in Section 3.8.1$_{148}$).

By comparing the definitions of the upper accessibility graph $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ and the critical graph $\mathcal{G}^c(\overline{\mathrm{T}})$ for an upper transition operator $\overline{\mathrm{T}}$, we see that the strongly connected components of $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ have to be unions of strongly connected components of $\mathcal{G}^c(\overline{\mathrm{T}})$. It is also not too difficult to see that the final classes of $\overline{\mathcal{G}}(\overline{\mathrm{T}})$ and the final classes of $\mathcal{G}^c(\overline{\mathrm{T}})$ are the same. This is exactly what allows us to check for top class regularity using the (usually much) cruder upper accessibility graph.

If the convex closed set of transition matrices $\mathcal{T}$ corresponding with $\overline{\mathrm{T}}$ is given explicitly in terms of a finite set of extreme points, then the calculation of the critical graph might be preferred over the calculation of the accessibility graph. However, if no finite set of extreme points is given, a vertex enumeration step is required (assuming that, unlike in Example 138$_{149}$, $\mathcal{T}$ has a finite number of extreme points). As it is provable that any algorithm based on vertex enumeration cannot have polynomial time complexity, the algorithm given by Akian and Gaubert becomes computationally intractable in this case. This is where our algorithm stands out. The reason it does, is because it works directly with the upper transition operator, and drops extra eigen space information that is not needed when checking for ergodicity.

3.11 The eigenvector approach to imprecise Markov chains

In the classical analysis of precise Markov chains, convergence results are usually reported in terms of eigenvalues of eigenvectors. A stationary distribution is, for example, a left eigenvector belonging to eigenvalue 1 of the transition matrix. For imprecise Markov chains, this approach will not be very fruitful in general. This is mainly because the linearity of the operator is lost. For imprecise Markov chains with a two-dimensional state space $\mathcal{X} = \{x, y\}$, a gamble can always be written as constant gamble plus a non-constant gamble and this sum can be pulled apart using constant additivity as will become clearer later on. In this section, we focus on such two-dimensional imprecise Markov chains.

In the binary case, the eigenequation is:

$$\overline{\mathrm{T}}\xi = \lambda\xi \Leftrightarrow \begin{cases} \overline{\mathrm{P}}(\xi|x) = \lambda\xi(x), \\ \overline{\mathrm{P}}(\xi|y) = \lambda\xi(y). \end{cases}$$

which can be written, using P4$_{65}$, as

$$\begin{cases} \xi(x) + \overline{\mathrm{P}}\left([\xi(y) - \xi(x)]I_{\{y\}}|x\right) = \lambda\xi(x), \\ \xi(y) + \overline{\mathrm{P}}\left([\xi(x) - \xi(y)]I_{\{x\}}|y\right) = \lambda\xi(y). \end{cases}$$

We can continue from here using non-negative homogeneity. However then we need to know the sign of $\xi(x) - \xi(y)$. Therefore, the problem splits in two; first we assume that $\xi(x) \geq \xi(y)$, and we call the corresponding solution ξ_x. Afterwards we will find the solution for $\xi(x) \leq \xi(y)$ and denote this solution by ξ_y.

If $\xi(x) \geq \xi(y)$, then we infer using P3$_{65}$ and the conjugacy relation $\overline{\mathrm{P}}(-f) = -\underline{\mathrm{P}}(f)$ that

$$\begin{cases} \xi(x) + [\xi(y) - \xi(x)]\,\underline{\mathrm{P}}\left(I_{\{y\}}|x\right) = \lambda\xi(x) \\ \xi(y) + [\xi(x) - \xi(y)]\,\overline{\mathrm{P}}\left(I_{\{x\}}|y\right) = \lambda\xi(y). \end{cases}$$

In matrix form, this equation looks like[11]

$$\underbrace{\begin{pmatrix} \lambda - 1 + \underline{\mathrm{P}}\left(I_{\{y\}}|x\right) & -\underline{\mathrm{P}}\left(I_{\{y\}}|x\right) \\ -\overline{\mathrm{P}}\left(I_{\{x\}}|y\right) & \lambda - 1 + \overline{\mathrm{P}}\left(I_{\{x\}}|y\right) \end{pmatrix}}_{=:(\lambda\mathbb{1}_2 - \underline{M}_{xy})} \underbrace{\begin{pmatrix} \xi(x) \\ \xi(y) \end{pmatrix}}_{\xi} = \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{0}.$$

This means that the non-linear eigenvalue problem, has been transformed to a linear one, with the extra constraint that the eigenfunctions ξ must satisfy $\xi(x) \geq \xi(y)$. The eigenvalues of $\underline{M}_{xy}$ are given by the zeros of its characteristic equation:

$$\begin{aligned} 0 = &(\lambda - 1)^2 + \left[\overline{\mathrm{P}}\left(I_{\{x\}}|y\right) + \underline{\mathrm{P}}\left(I_{\{y\}}|x\right)\right](\lambda - 1) \\ &+ \overline{\mathrm{P}}\left(I_{\{x\}}|y\right)\underline{\mathrm{P}}\left(I_{\{y\}}|x\right) - \overline{\mathrm{P}}\left(I_{\{x\}}|y\right)\underline{\mathrm{P}}\left(I_{\{y\}}|x\right), \end{aligned}$$

[11] here $\mathbb{1}_2$ stands for the two dimensional identity matrix.

or,

$$0 = \left(\lambda - \left[1 - \underline{\mathrm{P}}\left(I_{\{y\}}|x\right) - \overline{\mathrm{P}}\left(I_{\{x\}}|y\right)\right]\right)(\lambda - 1)$$
$$0 = \left(\lambda - \left[\overline{\mathrm{P}}\left(I_{\{x\}}|x\right) - \overline{\mathrm{P}}\left(I_{\{x\}}|y\right)\right]\right)(\lambda - 1).$$

This means that there are two solutions for $\xi(x) \geq \xi(y)$. The first solution, which we denote by ξ_1, is the eigenfunction belonging to the eigenvalue $\lambda = 1$. Solving $\underline{M}_{xy}(1) = 0$ shows us that ξ_1 must be a constant gamble. The first eigenvalue/eigenfunction pair is thus given by

$$\overline{\mathrm{T}} I_{\mathcal{X}} = I_{\mathcal{X}}. \tag{3.39}$$

This equation is fairly trivial as it can be derived from constant additivity.

The second solution ξ_x, is the eigenfunction belonging to the eigenvalue $\lambda_x := \overline{\mathrm{P}}\left(I_x|x\right) - \overline{\mathrm{P}}\left(I_x|y\right)$. Besides satisfying $\xi_x(x) \geq \xi_x(y)$, this solution has to satisfy

$$0 = \overline{\mathrm{P}}\left(I_{\{x\}}|y\right)\xi_x(x) + \underline{\mathrm{P}}\left(I_{\{y\}}|x\right)\xi_x(y).$$

Any non-negative multiple of $I_{\{x\}}\underline{\mathrm{P}}\left(I_{\{y\}}|x\right) - I_{\{y\}}\overline{\mathrm{P}}\left(I_{\{x\}}|y\right)$ satisfies both constraints. The eigenequation therefore becomes

$$\overline{\mathrm{T}}\underbrace{\left(I_{\{x\}}\underline{\mathrm{P}}\left(I_{\{y\}}|x\right) - I_{\{y\}}\overline{\mathrm{P}}\left(I_{\{x\}}|y\right)\right)}_{\xi_x} = \underbrace{\left[\overline{\mathrm{P}}\left(I_{\{x\}}|x\right) - \overline{\mathrm{P}}\left(I_{\{x\}}|y\right)\right]}_{\lambda_x}\ \xi_x. \tag{3.40}$$

In a similar manner, we can derive the eigenvalues and functions in the case that $\xi(x) \leq \xi(y)$. If we do so, then we see that $I_{\mathcal{X}}$ is again the eigen gamble belonging to eigenvalue 1. The other eigenvalue λ_y is then equal to $\lambda_y := \overline{\mathrm{P}}\left(I_{\{y\}}|y\right) - \overline{\mathrm{P}}\left(I_{\{y\}}|x\right)$ and the corresponding eigenfunction ξ_y is given by $\xi_y := -I_{\{x\}}\overline{\mathrm{P}}\left(I_{\{y\}}|x\right) + I_{\{y\}}\underline{\mathrm{P}}\left(I_{\{x\}}|y\right)$ and the eigenequation becomes

$$\overline{\mathrm{T}}\underbrace{\left(-I_{\{x\}}\overline{\mathrm{P}}\left(I_{\{y\}}|x\right) + I_{\{y\}}\underline{\mathrm{P}}\left(I_{\{x\}}|y\right)\right)}_{\xi_y} = \underbrace{\left[\overline{\mathrm{P}}\left(I_{\{y\}}|y\right) - \overline{\mathrm{P}}\left(I_{\{y\}}|x\right)\right]}_{\lambda_y}\ \xi_x. \tag{3.41}$$

Unlike the linear case, where the number of eigenvalues corresponds to the dimension of the problem, we now have three different eigenvalues and corresponding eigenfunctions.

If we apply the upper transition operator $\overline{\mathrm{T}}$ repeatedly to ξ_x, and we assume that neither $\xi_x = 0$ nor $\xi_y = 0$, then we can prove, using complete induction, that

$$\overline{\mathrm{T}}^n \xi_x = \begin{cases} \lambda_x^n \xi_x & \lambda_x \geq 0, \\ \nu_x \left(\frac{1-(\lambda_x\lambda_y)^{\lfloor \frac{n}{2} \rfloor}}{1-\lambda_x\lambda_y}(1-\lambda_x) - 1 \right) + (\lambda_x\lambda_y)^{\lfloor \frac{n}{2} \rfloor + 1} \left(\nu_x + \lambda_x^{n \bmod 2} \xi_x \right) & \\ & \lambda_x < 0 \text{ and } \lambda_y < 0 \text{ and } 1 < n, \\ -\nu_x\lambda_x(1-\lambda_y^{n-1}) + \lambda_x\lambda_y^{n-1}\xi_x & \text{otherwise.} \end{cases}$$

Here we have used that

$$-\xi_x = \nu_x + \frac{1-\lambda_x}{1-\lambda_y}\xi_y, \quad \text{with } \nu_x := \overline{\mathrm{P}}(I_{\{y\}}|x)\overline{\mathrm{P}}(I_{\{x\}}|y) - \underline{\mathrm{P}}(I_{\{y\}}|x)\underline{\mathrm{P}}(I_{\{x\}}|y),$$
$$-\xi_y = \nu_y + \frac{1-\lambda_y}{1-\lambda_x}\xi_x, \quad \text{with } \nu_y := \nu_x\frac{1-\lambda_y}{1-\lambda_x}.$$

What makes the two dimensional case especially attractive is that it is possible, whenever ξ_x and ξ_y are nonzero, to express every gamble f as the sum of a constant gamble and a non-negative multiple of one of the eigenvectors ξ_x or ξ_y:

$$f = \begin{cases} aI_{\mathcal{X}} + b\xi_x & f(x) \geq f(y) \\ aI_{\mathcal{X}} + b\xi_y & f(x) < f(y) \end{cases}$$

where $a \in \mathbb{R}$ and $b \in \mathbb{R}_{\geq 0}$. This means that we can calculate any power of $\overline{\mathrm{T}}$, applied to f as

$$\overline{\mathrm{T}}^n f = \begin{cases} a + b\overline{\mathrm{T}}^n\xi_x & f(x) \geq f(y) \\ a + b\overline{\mathrm{T}}^n\xi_y & f(x) < f(y) \end{cases}$$

We know that $|\lambda| \leq 1$ (otherwise, the transition operator would be expansive), so there is only one fixed point (up to a nonnegative multiple) if both $|\lambda_x| < 1$ and $|\lambda_y| < 1$. This fixed point will be a multiple of $\xi_1 := I_{\mathcal{X}}$, confirming Proposition 133$_{145}$ and Theorem 132$_{144}$. Observe that it is extremely easy to find the limit value of $\overline{\mathrm{T}}^n f$: assuming for example that $f(x) \geq f(y)$, we get that

$$\lim_{n\to\infty} \overline{\mathrm{T}}^n f = \begin{cases} a & \lambda_x \geq 0, \\ a - b\nu_x\lambda_x & \lambda_x < 0 \text{ and } \lambda_y \geq 0, \\ a - b\nu_x\lambda_x\frac{1-\lambda_y}{1-\lambda_x\lambda_y} & \lambda_x < 0 \text{ and } \lambda_y < 0. \end{cases}$$

CHAPTER **4**

Imprecise Markov trees

The last twenty years have witnessed a rapid growth of **graphical models** in the fields of artificial intelligence and statistics. These models combine graphs and probability to address complex multivariate problems in a variety of domains, such as medicine, finance, risk analysis, defence, and environment, to name just a few.

Much has been done also on the front of imprecise probability. In particular, **credal nets** [8] have been and still are the subject of intense research. A credal net creates a global model of a domain by combining local uncertainty models using some notion of independence, and then uses this to do inference. The local models represent uncertainty by closed convex sets of probabilities, also called **credal sets**, and the notion of independence used with credal nets in the vast majority of cases has been that of **strong independence** (with some exceptions in [5]).

An alternative and attractive approach to expressing independence that is not committed to the sensitivity analysis interpretation is offered by **epistemic irrelevance** (Section 1.5.168).

The question we address in this chapter should be clear: can we define credal nets based on epistemic irrelevance, and moreover create an exact algorithm to perform efficient inferences with them? We give a fully positive answer to this question in the special case that (i) the graph under consideration is a directed tree, and (ii) the related variables assume finitely many values. The intuitions that showed us the way towards this result originated in previous work on imprecise probability trees (see [12] and Chapter 2[80]) and imprecise Markov chains (see [14] and Chapter 3[122]) and lead to our paper on imprecise Markov trees [13], of which this chapter is a reflection.

In Section 4.1, we discuss preliminary graph-theoretic notions, and define the local uncertainty models that will be used at each node of a tree. These models are formalised through the language of **coherent lower previsions** (see Section 1.4.3[60]). We discuss how such local models will give rise to a global uncertainty model, which plays the same role as the joint mass function built by the chain rule in a Bayesian net. Based on the global model, we state the Markov Condition that defines the imprecise-probability interpretation of our credal trees. As was the case in the previous chapter, this Markov Condition involves epistemic irrelevance rather than strong independence.

In Section 4.3[168], we turn to the problem of constructing the most conservative global model based only on the local models in the tree and our Markov Condition. We show that this task can be achieved by a recursive construction that proceeds from the leaves to the root of the tree using two operations: the **independent natural extension** discussed in Section 1.5.1[68] and studied at length in [18, 19], and the **marginal extension**, defined in Theorem 30[51], and studied in great detail in [41, 62]. We also show that all uncertainty models we consider, the local ones as well as the global ones that we create, satisfy a consistency criterion that generalises (and is based on the same ideas as) the usual consistency criterion in Bayesian nets: they are (separately and jointly) **coherent** as in Lemma 29[51] [39, 40, 62, 69]. This is an important rationality requirement.

We briefly comment on some of the graphical separation criteria induced by epistemic irrelevance in Section 4.4[173]. We then go on to develop and justify an algorithm for making inferences on credal trees under epistemic irrelevance in Section 4.5[174]. The algorithm is used to **update** the tree: it computes posterior beliefs about a **target** variable in the tree conditional on the observation of other variables, which are called **instantiated**, meaning that their value is determined. It can in particular be used for treating the model as an expert system.

Our algorithm is based on message passing, as are the traditional algorithms that have been developed for precise graphical models. It has some remarkable properties: (i) it works in time essentially linear in the size of the tree; (ii) it natively computes posterior lower and upper **previsions** (or expectations) rather than probabilities; (iii) it is the first algorithm developed for credal nets that exclusively uses the formalism of coherent lower previsions; and (iv) we show that, under very mild conditions, using the tree for updating beliefs cannot lead to inferences that are inconsistent with the local models we have started from, nor with one another.

We give a step-by-step example of the way inferences can be drawn using our algorithm in Section 4.6_{182}. We also comment there on the intriguing relationship between the failure of certain classical separation properties in our framework, and the phenomenon of dilation [30, 50].

4.1 Credal trees under epistemic irrelevance

4.1.1 Basic notions and notation

We consider a rooted and directed discrete tree with finite width and depth. As the graphical structure is a tree, like in Chapter 2_{80}, a lot of the notation can be reused in the chapter. We call $\mathbb{S}$ the set of its nodes s, and we denote the **root**, or initial node, by $\square$. As will become clearer later on, the nodes in an imprecise Markov chain contain random variables, which is in contradistinction with imprecise probability trees, where the nodes contain situations. For any node s, we denote the set of its parents by $\mathrm{pa}(s)$. Of course, $\mathrm{pa}(\square) = \emptyset$, and for $s \neq \square$ we have that $\mathrm{pa}(s) = \{\mathrm{mo}(s)\}$ where $\mathrm{mo}(s)$ is the **mother node** of s. Also, for each node s, we denote the set of its **children** by $\mathrm{ch}(s)$, and the set of its **siblings** by $\mathrm{sib}(s)$. Clearly, $\mathrm{sib}(\square) = \emptyset$, and if $s \neq \square$ then $\mathrm{sib}(s) = \mathrm{ch}(\mathrm{mo}(s)) \setminus \{s\}$. If $\mathrm{ch}(s) = \emptyset$, then we call s a **leaf**, or **terminal node**. We denote by $\mathbb{S}^{\lozenge} := \{s \in \mathbb{S} : \mathrm{ch}(s) \neq \emptyset\}$ the set of all non-terminal nodes.

For nodes s and t, we write $s \sqsubseteq t$ if s **precedes** t, i.e., if there is a directed segment in the tree from s to t. The relation $\sqsubseteq$ is a special partial order on the set $\mathbb{S}$. $\mathrm{ancest}(s) := \{t \in \mathbb{S} : t \sqsubset s\}$ denotes the chain of **ancestors** of s, and $\mathrm{desc}(s) := \{t \in \mathbb{S} : s \sqsubset t\}$ its set of **descendants**. Here $s \sqsubset t$ means that $s \sqsubseteq t$ and $s \neq t$. We also use the notation $\downarrow s := \mathrm{ancest}(s) \cup \{s\}$ for the chain (segment) connecting $\square$ and s, and $\uparrow s := \mathrm{desc}(s) \cup \{s\}$ for the subtree with root s. Similarly, we let $\downarrow S := \bigcup\{\downarrow s : s \in S\}$ and $\uparrow S := \bigcup\{\uparrow s : s \in S\}$ for any subset $S \subseteq \mathbb{S}$.

With each node s of the tree, there is associated a variable X_s assuming values in a non-empty finite set $\mathcal{X}_s$. We extend this notation to more complicated situations as follows. If S is any subset of $\mathbb{S}$, then we denote by X_S the tuple of variables whose components are the X_s for all $s \in S$. This new joint variable assumes values in the finite set $\mathcal{X}_S = \times_{s \in S}\mathcal{X}_s$, and the corresponding set of gambles is denoted by $\mathcal{L}(\mathcal{X}_S)$.[1] Generic elements of $\mathcal{X}_s$ are denoted by x_s or z_s. Similarly for x_S and z_S in $\mathcal{X}_S$. Also, if we mention a tuple z_S, then for any $t \in S$, the corresponding element in the tuple will be denoted by z_t. We assume all variables in the tree to be **logically independent**, meaning that the variable X_S may assume **all** values in $\mathcal{X}_S$, for all $\emptyset \subseteq S \subseteq \mathbb{S}$. We also find it convenient to identify a gamble on $\mathcal{X}_S$ with its cylindrical extension to $\mathcal{X}_U$, where $S \subseteq U \subseteq \mathbb{S}$.

Throughout this chapter, we consider (conditional) lower previsions as models for a subject's beliefs about the values that certain variables in the tree may assume (see Section $1.4.3_{60}$). We use a systematic notation for such (conditional) lower previsions. Let $I, O \subseteq \mathbb{S}$ be **disjoint** sets of nodes with $O \neq \emptyset$, then we generically[2] denote by $\underline{V}_O(\cdot|X_I)$ a **conditional lower prevision** (see Definition 51_{66}), defined on the set of gambles $\mathcal{L}(\mathcal{X}_{I\cup O})$.[3] For every gamble f on $\mathcal{X}_{I\cup O}$ and every $x_I \in \mathcal{X}_I$, $\underline{V}_O(f|x_I)$ is the lower prevision (or lower expectation, or our subject's supremum buying price) for/of the gamble f, conditional on the event that $X_I = x_I$. We interpret $\underline{V}_O(f|X_I)$ as a real-valued map (gamble) on $\mathcal{X}_I$ that assumes the value $\underline{V}_O(f|x_I)$ in the element x_I of $\mathcal{X}_I$. The conjugate **conditional upper prevision** $\overline{V}_O(\cdot|X_I)$ is defined on $\mathcal{L}(\mathcal{X}_{I\cup O})$ by $\overline{V}_O(f|X_I) := -\underline{V}_O(-f|X_I)$ for all gambles f on $\mathcal{X}_{I\cup O}$. We will always implicitly assume that all conditional models $\underline{V}_O(\cdot|X_I)$ we use are **separately coherent**, meaning that they satisfy properties $\mathrm{P1}_{65}$, $\mathrm{P2}_{65}$ and $\mathrm{P3}_{65}$ of Proposition 48_{64} and that $\underline{V}_O(f|x_I) = \underline{V}_O(f(x_I,\cdot)|x_I)$. Hereafter, we will frequently introduce conditional lower previsions of the type $\underline{V}_O(\cdot|X_I)$ as if they are defined on $\mathcal{L}(\mathcal{X}_O)$, simply because that is a very natural thing to do: such a conditional lower prevision is usually interpreted as representing

[1] For any subset S of $\mathbb{S}$, $\mathcal{X}_S$ is defined formally as the set of all maps x_S of S to $\bigcup_{s\in S}\mathcal{X}_s$, such that $x_S(s) = x_s \in \mathcal{X}_s$ for all $s \in S$. So when $S = \emptyset$, the empty product $\mathcal{X}_\emptyset$ is defined as the set of all maps from $\emptyset$ to $\emptyset$, which is a singleton. The corresponding variable $X_\emptyset$ can then only assume this single value, so there is no uncertainty about it. $\mathcal{L}(\mathcal{X}_\emptyset)$ can be identified with the set $\mathbb{R}$ of real numbers.

[2] Besides the letter V, we will also use the letters P, Q and R.

[3] In keeping with the observation in footnote 1, we also allow $I = \emptyset$, which means conditioning on the variable $X_I = X_\emptyset$, which can only assume one single value. This means that $\underline{V}_O(\cdot|X_\emptyset) =: \underline{V}_O$ effectively becomes an unconditional lower prevision on $\mathcal{L}(\mathcal{X}_{O\cup\emptyset}) = \mathcal{L}(\mathcal{X}_O)$. This is a very useful device that allows us to use the same generic notation for both conditional and unconditional lower previsions.

beliefs about the variable X_O, conditional on values of the variable X_I.

As soon as we consider a number of such conditional lower previsions $\underline{\mathrm{V}}_{O_k}(\cdot|X_{I_k})$, $k = 1, \dots, n$, they should satisfy more stringent consistency criteria than that each of them should be separately coherent: they should also be consistent with one another in the sense of Walley's **(joint) coherence**. For more details about this much more involved type of coherence, we refer to [39, 40, 62] and Definition 29_{51}.

4.1.2 Local uncertainty models

We now add a **local uncertainty model** to each of the nodes s. If s is not the root node, i.e. has a mother $\mathrm{mo}(s)$, then this local model is a (separately coherent) conditional lower prevision $\underline{\mathrm{Q}}_s(\cdot|X_{\mathrm{mo}(s)})$ on $\mathcal{L}(\mathcal{X}_s)$: for each possible value $z_{\mathrm{mo}(s)}$ of the variable $X_{\mathrm{mo}(s)}$ associated with its mother $\mathrm{mo}(s)$, we have a coherent lower prevision $\underline{\mathrm{Q}}_s(\cdot|z_{\mathrm{mo}(s)})$ for the value of X_s, conditional on $X_{\mathrm{mo}(s)} = z_{\mathrm{mo}(s)}$. In the root, we have an unconditional local uncertainty model $\underline{\mathrm{Q}}_\square$ for the value of $X_\square$; $\underline{\mathrm{Q}}_\square$ is a (separately) coherent lower prevision on $\mathcal{L}(\mathcal{X}_\square)$. We use the common generic notation $\underline{\mathrm{Q}}_s(\cdot|X_{\mathrm{pa}(s)})$ for all these local models.

4.1.3 Global uncertainty models

We intend to show in Section 4.3_{168} how all these local models $\underline{\mathrm{Q}}_s(\cdot|X_{\mathrm{mo}(s)})$ can be combined into **global uncertainty models**. We generically denote such global models using the letter P. More specifically, we want to end up with an unconditional joint lower prevision $\underline{\mathrm{P}} := \underline{\mathrm{P}}_{\uparrow\square} = \underline{\mathrm{P}}_{\mathbb{S}}$ on $\mathcal{L}(\mathcal{X}_{\mathbb{S}})$ for all variables in the tree, as well as conditional lower previsions $\underline{\mathrm{P}}_{\uparrow s}(\cdot|X_{\mathrm{mo}(s)})$ on $\mathcal{L}(\mathcal{X}_{\uparrow s})$ for all non-initial nodes s, and $\underline{\mathrm{P}}_{\uparrow\mathrm{ch}(t)}(\cdot|X_t)$ on $\mathcal{L}(\mathcal{X}_{\uparrow\mathrm{ch}(t)})$ for all non-terminal nodes t.

Ideally, we want these global (conditional) lower previsions (i) to be compatible with the local assessments $\underline{\mathrm{Q}}_s(\cdot|X_{\mathrm{mo}(s)})$, $s \in \mathbb{S}$, (ii) to be coherent with one another, and (iii) to reflect the conditional irrelevancies (or Markov-type conditions) that we want the graphical structure of the tree to encode. In addition, we want them (iv) to be as conservative (small) as possible.

In this list, the only item that needs more explanation concerns the Markov-type conditions that the tree structure encodes. This is what we turn to now.

4.1.4 The interpretation of the graphical model

In classical Bayesian nets, the graphical structure is taken to represent the following assessments: for any node s, conditional on its parent variables, its non-parent non-descendant variables are epistemically irrelevant to it (and therefore also independent).

In the present context, we assume that the tree structure embodies the following conditional irrelevance assessment, which turns out to be equivalent with the conditional independence assessment above in the special case of a Bayesian tree.

CI. Consider any node s in the tree, any subset S of its set of children $\mathrm{ch}(s)$, and the set[4] $\perp(S) := \bigcap_{c\in S} \perp(c)$ of their common non-parent non-descendants. Then **conditional on the mother variable X_s, the non-parent non-descendant variables $X_{\perp(S)}$ are assumed to be epistemically irrelevant to the variables $X_{\uparrow S}$ associated with the children in S and their descendants**:

$$X_{\perp(S)} \,\mathrm{EI}\, X_{\uparrow S} \,|\, X_s.$$

This interpretation turns the tree into a **credal tree under epistemic irrelevance**, and we also introduce the term **imprecise Markov tree** (IMT) for it. For the global models we are considering here, CI has the following consequences. It implies that for all $s \in \mathbb{S}^{\Diamond}$, all non-empty $S \subseteq \mathrm{ch}(s)$ and all $I \subseteq \perp(S)$, we can infer from $\underline{P}_{\uparrow S}(\cdot|X_s)$ a model $\underline{P}_{\uparrow S}(\cdot|X_{\{s\}\cup I})$, where for all $z_{\{s\}\cup I} \in \mathcal{X}_{\{s\}\cup I}$, with obvious notations:[5]

$$\underline{P}_{\uparrow S}(f|z_{\{s\}\cup I}) := \underline{P}_{\uparrow S}(f(\cdot, z_I)|z_s) \text{ for all gambles } f \text{ in } \mathcal{L}(\mathcal{X}_{\uparrow S\cup I}), \tag{4.1}$$

where $f(\cdot, z_I)$ denotes a partial map of f, defined on $\mathcal{X}_{\uparrow S}$.

We discuss some of the separation properties that accompany this interpretation in Section 4.4_{173}. For now, we focus on two immediate consequences that will help us go from local to global models in Section 4.3_{168}.

First, consider some node s. Then CI tells us that for any two children $c_1, c_2 \in \mathrm{ch}(s)$ of s, the variable $X_{\uparrow c_1}$ is epistemically irrelevant to the variable $X_{\uparrow c_2}$, conditional on X_s.

[4] Here the set $\perp(c)$ of non-parents non-descendants of the node c is defined by $\perp(c) := \mathbb{S} \setminus (\mathrm{pa}(c) \cup \mathrm{desc}(c))$.

[5] For leaves s, the corresponding irrelevance condition is trivial, as the set $\mathrm{ch}(s)$ of children of s is empty.

$$\begin{tikzcd}[column sep=small] & X_s \arrow[dl] \arrow[d] \arrow[dr] & \\ X_{\uparrow c_1} & \cdots & X_{\uparrow c_2} \end{tikzcd}$$

It even tells us that for any two disjoint non-empty sets $S_1 \subseteq \mathrm{ch}(s)$ and $S_2 \subseteq \mathrm{ch}(s)$ of children of s, the variable $X_{\uparrow S_1}$ is epistemically irrelevant to $X_{\uparrow S_2}$, conditional on X_s. We conclude that, conditional on a node, all its children c (and the variables associated with their sub-trees $\uparrow c$) are **epistemically independent** (see Section 1.5.1$_{68}$), in the specific sense to be discussed in the next section.

Next, consider some non-terminal node s different from $\square$, and its mother variable $X_{\mathrm{mo}\,s}$. We infer from CI that this mother variable $X_{\mathrm{mo}\,s}$ is epistemically irrelevant to the variable $X_{\uparrow \mathrm{ch}(s)}$ conditional on X_s:

$$\begin{tikzcd}[column sep=small] & X_{\mathrm{mo}(s)} \arrow[d] & \\ & X_s \arrow[dl] \arrow[d] \arrow[dr] & \\ X_{\uparrow c_1} & \cdots & X_{\uparrow c_n} \end{tikzcd} \quad \text{or equivalently,} \quad \begin{tikzcd} X_{\mathrm{mo}(s)} \arrow[d] \\ X_s \arrow[d] \\ X_{\uparrow \mathrm{ch}(s)} \end{tikzcd}$$

4.2 Factorising lower previsions

The following notion of factorisation is intimately linked with that of an independent product (see [17, Theorem 24] and Section 1.5.1$_{68}$). It will also play a crucial part in our development of an algorithm for updating an imprecise Markov tree in Section 4.5$_{174}$.

Definition 142: *We call a (separately) coherent lower prevision $\underline{\mathrm{P}}_N$ on $\mathcal{L}(\mathcal{X}_N)$ **strongly factorising** if for all disjoint proper subsets O and I of N, all $g \in \mathcal{L}(\mathcal{X}_O)$ and all non-negative $f \in \mathcal{L}(\mathcal{X}_I)$, $\underline{\mathrm{P}}_N(fg) = \underline{\mathrm{P}}_N(f\underline{\mathrm{P}}_N(g))$.*

As an important example, the so-called **strong product** $\boxtimes_{n\in N}\underline{\mathrm{P}}_n$ [8] of the marginal lower previsions $\underline{\mathrm{P}}_n$ is factorising [19].[6]

[6]This type of independent product comes to the fore in a study of credal nets under strong independence.

As a consequence of the separate coherence of the joint lower prevision $\underline{P}_N$, the right-hand side of the equality in this definition can be rewritten as:

$$\underline{P}_N(f\underline{P}_N(g)) = \begin{cases} \underline{P}_N(f)\underline{P}_N(g) & \text{if } \underline{P}_N(g) \geq 0 \\ \overline{P}_N(f)\underline{P}_N(g) & \text{if } \underline{P}_N(g) \leq 0 \end{cases}.$$

If $f_i \in \mathcal{L}_{\geq 0}(\mathcal{X}_i)$ for $i \in I$ then this equation implies that

$$\underline{P}_N(\underline{P}_N(g\prod_{i\in I})f_i) = \begin{cases} \underline{P}_o(g)\prod_{i\in I}\underline{P}_i(f_i) & \text{if } \underline{P}_o(g) \geq 0 \\ \underline{P}_o(g)\prod_{i\in I}\overline{P}_i(f_i) & \text{if } \underline{P}_o(g) \leq 0, \end{cases}$$

which explains where the term 'factorising' comes from. In particular, for any (separately) coherent factorising joint lower prevision $\underline{P}_N$, we see that for any non-empty subset I of N:

$$\underline{P}_N(\times_{i\in I}A_i) = \prod_{i\in I}\underline{P}_N(A_i) \text{ and } \overline{P}_N(\times_{i\in I}A_i) = \prod_{i\in I}\overline{P}_N(A_i), \tag{4.2}$$

where $A_i \subseteq \mathcal{X}_i$ for all $i \in I$.

The independent natural extension has very interesting and non-trivial **marginalisation and associativity properties** (Proposition 60_{72}). Consider any non-empty subset R of N, then the independent natural extension satisfies

$$\otimes_{n\in N}\mathcal{A}_{X_n} = \left(\otimes_{r\in R}\mathcal{A}_{X_r}\right)\otimes\left(\otimes_{n\in N\setminus R}\mathcal{A}_{X_n}\right) \quad \text{and}$$
$$\operatorname{marg}_R\left(\otimes_{n\in N}\mathcal{A}_{X_n}\right) = \otimes_{r\in R}\mathcal{A}_{X_r}.$$

The corresponding lower prevision for any gamble $f \in \mathcal{L}(\mathcal{X}_N)$ will be given by

$$(\otimes_{n\in N}\underline{P}_n)(f) := \left\{\alpha \in \mathbb{R}: f - \alpha \in \otimes_{n\in N}\mathcal{A}_{X_n}\right\} \tag{4.3}$$

and they satisfy (see also [19] and Proposition 62_{73})

$$\left(\otimes_{r\in R}\underline{P}_r\right)(g) = \left(\otimes_{n\in N}\underline{P}_n\right)(g) \text{ for all gambles } g \text{ on } \mathcal{X}_R. \tag{4.4}$$

Moreover, for any partition N_1 and N_2 of N, we have that

$$\otimes_{n\in N}\underline{P}_n = \left(\otimes_{n_1\in N_1}\underline{P}_{n_1}\right)\otimes\left(\otimes_{n_2\in N_2}\underline{P}_{n_2}\right), \tag{4.5}$$

so $\otimes_{n\in N}\underline{P}_n$ is the independent natural extension of its $\mathcal{X}_{N_1}$-marginal $\otimes_{n_1\in N_1}\underline{P}_{n_1}$ and its $\mathcal{X}_{N_2}$-marginal $\otimes_{n_2\in N_2}\underline{P}_{n_2}$.

If confronted with a finite set of marginal lower previsions, $\underline{P}_n$ on $\mathcal{X}_n$ with $n \in N$, then De Cooman, Miranda and Zaffalon proved [19, Theorem 24], that the independent natural extension $\otimes_{n \in N} \underline{P}_n$ is factorising.

Theorem 143: *Consider coherent lower previsions $\underline{P}_n$ on $\mathcal{L}(\mathcal{X}_n)$, $n \in N$. Then their independent natural extension $\otimes_{n \in N} \underline{P}_n$ is factorising.*

4.3 Constructing the most conservative joint

Let us show how to construct specific global models for the variables in the tree, and argue that these are the most conservative coherent models that extend the local models and express all conditional irrelevancies (4.1), encoded in the imprecise Markov tree. In Section 4.5$_{174}$, we will use these global models to construct and justify an algorithm for updating the imprecise Markov tree.

The crucial step lies in the recognition that any tree can be constructed recursively from the leaves up to the root, by using basic building blocks of the following type:

$$\begin{tikzcd}
& X_{\mathrm{mo}(s)} \arrow[d] & & \\
& X_s \arrow[dl] \arrow[d] \arrow[dr] \arrow[rr, dotted] & & \underline{Q}_s(\cdot|X_{\mathrm{mo}(s)}) \\
X_{\uparrow c_1} & X_{\uparrow c_2} & X_{\uparrow c_n} \arrow[r, dotted] & \underline{P}_{\uparrow c_k}(\cdot|X_s)
\end{tikzcd}$$

The global models are then also constructed recursively, following the same pattern. In what follows, we first derive the recursion equations for these global models in a heuristic manner. The real justification for using the global models thus derived is then given in Theorem 146$_{172}$.

Consider a node s and suppose that, in each of its children $c \in \mathrm{ch}(s)$, we already have a global conditional lower prevision $\underline{P}_{\uparrow c}(\cdot|X_s)$ on $\mathcal{L}(\mathcal{X}_{\uparrow c})$ [or equivalently, on $\mathcal{L}(\mathcal{X}_{\{s\} \cup \uparrow c})$].

Given that, conditional on X_s, the variables $X_{\uparrow c}$, $c \in \mathrm{ch}(s)$ are epistemically independent [see Section 4.1.4$_{165}$, condition CI], the discussion in Section 4.2$_{166}$ leads us to combine the 'marginals' $\underline{P}_{\uparrow c}(\cdot|X_s)$, $c \in \mathrm{ch}(s)$ into their point-wise smallest conditionally independent product (conditionally independent natural extension) $\otimes_{c \in \mathrm{ch}(s)} \underline{P}_{\uparrow c}(\cdot|X_s)$, which is a conditional lower prevision $\underline{P}_{\uparrow \mathrm{ch}(s)}(\cdot|X_s)$ on $\mathcal{L}(\mathcal{X}_{\uparrow \mathrm{ch}(s)})$ [or equivalently, on $\mathcal{L}(\mathcal{X}_{\uparrow s})$]:

$$\begin{tikzcd}
X_{\mathrm{mo}(s)} \arrow[d] & \\
X_s \arrow[r, dotted] \arrow[d] & \underline{Q}_s(\cdot|X_{\mathrm{mo}(s)}) \\
X_{\uparrow\mathrm{ch}(s)} \arrow[r, dotted] & \otimes_{c\in\mathrm{ch}(s)}\underline{P}_{\uparrow c}(\cdot|X_s) =: \underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_s)
\end{tikzcd}$$

Next, we need to combine the conditional models $\underline{Q}_s(\cdot|X_{\mathrm{mo}(s)})$ and $\underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_s)$ into a global conditional model about $X_{\uparrow s}$. Given that, conditional on X_s, the variable $X_{\mathrm{mo}\,s}$ is epistemically irrelevant to the variable $X_{\uparrow\mathrm{ch}(s)}$ [see Section 4.1.4165, condition CI], we expect $\underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_{\{\mathrm{mo}\,s,s\}})$ and $\underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_s)$ to coincide [this is a special instance of Equation (4.1)]. The most conservative (point-wise smallest) coherent way of combining the conditional lower previsions $\underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_{\{\mathrm{mo}\,s,s\}})$ and $\underline{Q}_s(\cdot|X_{\mathrm{mo}(s)})$ consists in taking their **marginal extension** $\underline{Q}_s(\underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_{\{\mathrm{mo}\,s,s\}})|X_{\mathrm{mo}(s)}) = \underline{Q}_s(\underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_s)|X_{\mathrm{mo}(s)})$; see [41, 62] and Theorem 3051 for more details. Graphically:

$$\begin{tikzcd}
X_{\mathrm{mo}(s)} \arrow[d] & \\
X_{\uparrow s} \arrow[r, dotted] & \underline{Q}_s(\underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_s)|X_{\mathrm{mo}(s)}) =: \underline{P}_{\uparrow s}(\cdot|X_{\mathrm{mo}(s)})
\end{tikzcd}$$

Summarising, and also accounting for the case $s = \square$, we can construct a global conditional lower prevision $\underline{P}_{\uparrow s}(\cdot|X_{\mathrm{mo}(s)})$ on $\mathcal{L}(\mathcal{X}_{\uparrow s})$ by backwards recursion:

$$\underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_s) := \otimes_{c\in\mathrm{ch}(s)}\underline{P}_{\uparrow c}(\cdot|X_s) \tag{4.6}$$

$$\underline{P}_{\uparrow s}(\cdot|X_{\mathrm{mo}(s)}) := \underline{Q}_s(\underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_s)|X_{\mathrm{mo}(s)}) = \underline{Q}_s(\otimes_{c\in\mathrm{ch}(s)}\underline{P}_{\uparrow c}(\cdot|X_s)|X_{\mathrm{mo}(s)}), \tag{4.7}$$

for all $s \in \mathbb{S}^{\lozenge}$. If we start with the 'boundary conditions'

$$\underline{P}_{\uparrow t}(\cdot|X_{\mathrm{mo}(t)}) := \underline{Q}_t(\cdot|X_{\mathrm{mo}(t)}) \text{ for all leaves } t, \tag{4.8}$$

then the recursion relations (4.6) and (4.7) eventually lead to the global joint model $\underline{P}_{\square} = \underline{P}_{\uparrow\square}(\cdot|X_{\mathrm{mo}(\square)})$, and to the global conditional models $\underline{P}_{\uparrow\mathrm{ch}(s)}(\cdot|X_s)$ for all non-terminal nodes s. For any subset $S \subseteq \mathrm{ch}(s)$, the global conditional model $\underline{P}_{\uparrow S}(\cdot|X_s)$ can then be defined simply as the restriction of the model $\underline{\mathrm{P}}_{\uparrow\mathrm{ch}(s)}(\cdot|X_s)$ on $\mathcal{L}(\mathcal{X}_{\uparrow\mathrm{ch}(s)})$ to the set $\mathcal{L}(\mathcal{X}_{\uparrow S})$:

$$\underline{\mathrm{P}}_{\uparrow S}(g|X_s) := \underline{\mathrm{P}}_{\uparrow\mathrm{ch}(s)}(g|X_s) \text{ for all gambles } g \text{ on } \mathcal{X}_{\uparrow S}. \tag{4.9}$$

It follows from the discussion leading to Equation (4.6) that, more directly [see Equation (4.4)],

$$\underline{\mathrm{P}}_{\uparrow S}(\cdot|X_s) = \otimes_{c\in S}\underline{\mathrm{P}}_{\uparrow c}(\cdot|X_s). \tag{4.10}$$

For easy reference, we will in what follows refer to this collection of global models as the **family of global models** $\mathcal{T}(\underline{P})$, so

$$\mathcal{T}(\underline{P}) := \{\underline{\mathrm{P}}\} \cup \left\{\underline{\mathrm{P}}_{\uparrow S}(\cdot|X_s) : s \in \mathbb{S}^{\lozenge} \text{ and non-empty } S \subseteq \mathrm{ch}(s)\right\}.$$

We end this section by discussing a number of interesting properties for the family of global models $\mathcal{T}(\underline{P})$ we can derive in this way. Let us call any real functional Φ on $\mathcal{L}$ **strictly positive** if $\Phi(I_{\{x\}}) > 0$ for all $x \in \mathcal{X}$. The proofs of the following two propositions can be found in [13].

Proposition 144: *If all the local models $\overline{\mathrm{Q}}_s(\cdot|X_{\mathrm{mo}(s)})$, $s \in \mathbb{S}$ are strictly positive, then so are all the global models in $\mathcal{T}(\underline{P})$.*

Proposition 145: *Consider any non-empty subset E of $\mathbb{S}$ and any $x_E \in \mathcal{X}_E$. If $\overline{\mathrm{P}}(\{x_E\}) > 0$ then also $\overline{\mathrm{P}}_{\uparrow c}(\{x_{E\cap\uparrow c}\}|x_e) > 0$ for all $e \in E$ and all $c \in \mathrm{ch}(e)$.*[7]

Before we formulate the most important result in this section (and arguably, in this chapter), we provide some motivation. Suppose we have some family of global models

$$\mathcal{T}(\underline{V}) := \{\underline{\mathrm{V}}\} \cup \left\{\underline{\mathrm{V}}_{\uparrow S}(\cdot|X_s) : s \in \mathbb{S}^{\lozenge} \text{ and non-empty } S \subseteq \mathrm{ch}(s)\right\}$$

associated with the tree. How do we express that such a family is compatible with the assessments encoded in the tree?

First of all, we require that our global models should extend the local models:

V1. For each $s \in \mathbb{S}$, $\underline{\mathrm{Q}}_s(\cdot|X_{\mathrm{mo}(s)})$ is the restriction of $\underline{\mathrm{V}}_{\uparrow s}(\cdot|X_{\mathrm{mo}(s)})$ to $\mathcal{L}(\mathcal{X}_s)$.

The second requirement is that our models should satisfy the rationality requirement of coherence:

[7] Observe that this holds trivially also if $E \cap \uparrow c = \emptyset$, because then $\mathcal{X}_{E\cap\uparrow c} = \mathcal{X}_{\emptyset}$ is a singleton [see footnote 1_{163}] whose upper probability is 1 by separate coherence.

V2. The (conditional) lower previsions in $\mathcal{T}(\underline{V})$ are jointly coherent.

The third requirement needs some preparatory explanation: the global models should reflect all epistemic irrelevancies encoded in the graphical structure of the tree. Naively, we would want condition (4.1) to be satisfied. The problem is that only the right-hand side in Equation (4.1), involving the model $\underline{\mathrm{V}}_{\uparrow S}(\cdot|X_s)$ is directly available to us. To get to the left-hand side involving the model $\underline{\mathrm{V}}_{\uparrow S}(\cdot|X_{\{s\}\cup I})$, one naive approach would be to 'condition the joint model $\underline{\mathrm{V}} = \underline{\mathrm{V}}_T$ on the variable $X_{\{s\}\cup I}$'. But we have seen in Proposition 53_{67} that given a joint model, coherence in general only determines the conditional models uniquely, provided that the **lower probability** of the conditioning event is non-zero. This is a fairly strong condition, and in what follows we would generally prefer to work with the much weaker condition that the **upper probability** of the conditioning event is non-zero.

Nevertheless, as soon as we realise that all we can reasonably require from our models is that they should be coherent, the right approach readily suggests itself: we should require that if we use the available models $\underline{\mathrm{V}}_{\uparrow S}(\cdot|X_s)$ to **define** the models $\underline{\mathrm{V}}_{\uparrow S}(\cdot|X_{\{s\}\cup I})$ through the epistemic irrelevance condition (4.1), then the result should still be coherent:

V3. If we define the conditional lower previsions $\underline{\mathrm{V}}_{\uparrow S}(\cdot|X_{\{s\}\cup R})$, $s \in \mathbb{S}^{\lozenge}$, $S \subseteq \mathrm{ch}(s)$ and $R \subseteq \perp(S)$ through the epistemic irrelevance requirements

$$\underline{\mathrm{V}}_{\uparrow S}(f|z_{\{s\}\cup R}) := \underline{\mathrm{V}}_{\uparrow S}(f(\cdot, z_R)|z_s) \text{ for all gambles } f \text{ in } \mathcal{L}(\mathcal{X}_{\uparrow S\cup R}),$$

then all these models together should be (jointly) coherent with all the available models in the family $\mathcal{T}(\underline{V})$.

And there is a final requirement, which guarantees that all inferences we make on the basis of our global models are as conservative as possible, and are therefore based on no other considerations than what is encoded in the tree:

V4. The models in the family $\mathcal{T}(\underline{V})$ are dominated (point-wise) by the corresponding models in all other families satisfying requirements TV1–TV3.

It turns out that the family of models $\mathcal{T}(\underline{P})$ we have been constructing above satisfy all four requirements.

Theorem 146: *If all local models $\overline{\mathrm{Q}}_s(\cdot|X_{\mathrm{mo}(s)})$ on $\mathcal{L}(\mathcal{X}_s)$, $s \in \mathbb{S}$ are strictly positive, then the family of global models $\mathcal{T}(\underline{P})$, obtained through Equations* (4.6)–(4.9)*, constitutes the point-wise smallest family of (conditional) lower previsions that satisfy* TV1$_{170}$–TV3*. It is therefore the unique family to also satisfy* TV4*. Finally, consider any non-empty set of nodes $E \subseteq \mathbb{S}$ and the corresponding conditional lower prevision derived by applying regular extension:*

$$\underline{\mathrm{R}}(f|x_E) := \max\left\{\mu \in \mathbb{R}: \underline{\mathrm{P}}_{\uparrow T}(I_{\{x_E\}}[f-\mu]) \geq 0\right\}$$

for all $f \in \mathcal{L}(\mathcal{X}_T)$ and all $x_E \in \mathcal{X}_E$. Then the conditional lower prevision $\underline{\mathrm{R}}(\cdot|X_E)$ is (jointly) coherent with the global models in the family $\mathcal{T}(\underline{P})$.

The last statement of this theorem guarantees that if we use regular extension to **update the tree** given evidence $X_E = x_E$, i.e., derive conditional models $\underline{\mathrm{R}}(\cdot|x_E)$ from the joint model $\underline{\mathrm{P}} = \underline{\mathrm{P}}_{\uparrow T}$, such inferences will always be coherent. This is of particular relevance for the discussion in Section 4.5$_{174}$, where we derive an efficient algorithm for updating the tree using regular extension. It implies in particular that our algorithm produces coherent inferences.

Proof: We will prove this theorem using the machinery of sets of acceptable gambles developed in the first chapter. We start by picking, for every node s, any tuple (one for every z in $\mathcal{X}_{\mathrm{mo}(s)}$) of coherent sets of acceptable gambles $\mathcal{A}_s\rfloor X_{\mathrm{mo}(s)}$ that is compatible with the local coherent conditional upper prevision $\overline{\mathrm{Q}}_s(\cdot|X_{\mathrm{mo}(s)})$.

The proper combination of these sets of acceptable gambles is done as explained, by applying marginal extension and independent natural extension in an iterative fashion:

$$\mathcal{A}_{\downarrow s}\rfloor X_{\mathrm{mo}(s)} = \begin{cases} \mathcal{A}_s\rfloor X_{\mathrm{mo}(s)} + \sum_{x_s \in \mathcal{X}_s} I_{\{z\}}\left(\otimes_{c \in \mathrm{ch}(s)} \mathcal{A}_{\downarrow c}\rfloor z\right) & \text{when } s \in \mathbb{S}^{\Diamond}, \\ \mathcal{A}_s\rfloor X_{\mathrm{mo}(s)} & \text{when } s \text{ is terminal.} \end{cases} \tag{4.11}$$

It follows immediately from the finite character of the tree, Properties 27$_{49}$ and 25$_{48}$, Theorem 30$_{51}$ and Lemma 59$_{71}$, that $\mathcal{A}_{\downarrow s}\rfloor X_{\mathrm{mo}(s)}$ is the smallest (jointly) coherent set of acceptable gambles that, by construction, encodes all necessary epistemic irrelevancies.

We moreover know from Propositions 144$_{170}$ and 145$_{170}$ that the joint set of almost acceptable gambles, $\mathcal{A}_{\downarrow\Box}$, will not contain a practically impossible event. This means that $\mathcal{A}_{\downarrow s}\rfloor X_{\mathrm{mo}(s)}$ is the result of properly updating and marginalising $\mathcal{A}_{\downarrow\Box}$.

Clearly, the resulting joint set of acceptable gambles $\mathcal{A}_{\downarrow\Box}$ does depend on the local sets of acceptable gambles we have chosen initially. However, by 'translating' (use

Theorem 112_{130} for Marginal Extension and the Proposition 62_{73} for the independent naturals extension) Equation (4.12) to lower previsions, we infer that

$$\underline{P}_{\uparrow s}(\cdot|X_{\mathrm{mo}(s)}) = \begin{cases} \underline{Q}_s(\otimes_{c\in\mathrm{ch}(s)}\underline{P}_{\uparrow c}(\cdot|X_s)|X_{\mathrm{pa}(s)}) & \text{when } s \in \mathbb{S}^{\lozenge}, \\ \underline{Q}_s(\cdot|X_{\mathrm{pa}(s)}) & \text{when } s \text{ is terminal.} \end{cases} \tag{4.12}$$

which are exactly Equations (4.7) and (4.8). More importantly, these equations do no longer depend on the local sets of acceptable gambles that were chosen initially. Hence, we may conclude that the global models $\underline{P}_{\uparrow s}(\cdot|X_{\mathrm{mo}(s)})$ satisfy V1$_{170}$–V3$_{171}$ and V4$_{171}$.

In the beginning of the proof, we were allowed to select any local set of acceptable gambles $\mathcal{A}_s$, compatible with the local lower prevision $\overline{Q}_s(\cdot|X_{\mathrm{pa}(s)})$. If we take not just any set, but the (unique) set of almost desirable gambles $\mathcal{R}_s$, then we infer from Proposition 53_{67} that $\underline{R}(\cdot|x_E)$ is the lower prevision that is compatible with the joint set of almost desirable gambles, updated on x_E, $\mathcal{R}_{\downarrow\square}\rfloor x_E$, whence $\underline{R}(\cdot|x_E)$ has to be coherent with the global models $\underline{P}_{\uparrow s}(\cdot|X_{\mathrm{pa}(s)})$, $s \in \mathbb{S}$ and $\underline{P}_{\uparrow \mathrm{ch}(s)}(\cdot|X_s)$. □

4.4 Some separation properties

Without going into too much detail, we would like to point out some of the more striking differences between the separation properties in imprecise Markov trees under epistemic irrelevance, and the more usual ones that are valid for Bayesian nets [47], which are also inherited from Bayesian nets by credal nets under strong independence [8].

It is clear from the interpretation of the graphical model described in Section 4.1.4$_{165}$ that we have the following simple separation results:

$$X_{i_1} \longrightarrow X_{i_2} \longrightarrow X_t \qquad\qquad X_{i_1} \longleftarrow X_{i_2} \longrightarrow X_t$$

where in both cases, X_{i_2} **separates** X_t **from** X_{i_1}: when the value of X_{i_2} is known, additional information about the value of X_{i_1} does not affect beliefs about the value of X_t. In this figure, between i_1 and i_2, and between i_2 and t, there may be other nodes, but the arrows along the path segment through these nodes should all point in the indicated directions. The underlying idea is that t is a (descendant of some) child c of i_2, and conditional on the mother i_2 of c, the non-parent non-descendant i_1 of c is epistemically irrelevant to c and all of its descendants.

On the other hand, and in contradistinction with what we are used to in Bayesian nets, we will not generally have separation in the following configuration:

$$X_{i_1} \longleftarrow X_{i_2} \longleftarrow X_t$$

where X_{i_2} **does not necessarily separate** X_t **from** X_{i_1}. We will come across a simple counterexample in Section 4.6182. Where does this difference with the case of Bayesian nets originate? It is clear from the reasoning above that X_{i_2} separates X_{i_1} from X_t: conditional on X_{i_2}, X_t is epistemically irrelevant to X_{i_1}. For precise probability models, irrelevance always implies symmetrical independence, and therefore this will imply that conditional on X_{i_2}, X_{i_1} is epistemically irrelevant to X_t as well. But for imprecise probability models no such symmetry is guaranteed [7], and we therefore cannot infer that, generally speaking, X_{i_2} will separate X_{i_1} from X_t. **As a general rule, we can only infer separation if the arrows point from the 'separating' variable X_{i_2} towards the 'target' variable X_t.**

4.5 An efficient algorithm for updating in an imprecise Markov tree

We now consider the case where we are interested in drawing inferences about the value of the variable X_t in some **target node** t, when we know the values x_E of the variables X_E in a set $E \subseteq \mathbb{S} \setminus \{t\}$ of **evidence nodes**; see for instance Fig 4.1179 on page 179.

4.5.1 The formulation of the problem

If we assume that the values of the remaining variables are **missing at random**, then we can do this by conditioning the joint $\underline{\mathrm{P}}$ obtained above on the available evidence '$X_E = x_E$'; see for instance [22, 70].

We will address this problem by updating the lower prevision $\underline{\mathrm{P}}$ to the lower prevision $\underline{\mathrm{R}}_t(\cdot|x_E)$ on $\mathcal{L}(\mathcal{X}_t)$ using **regular extension** (see Proposition 5367 and [62, Appendix J]):

$$\underline{\mathrm{R}}_t(g|x_E) = \max\left\{\mu \in \mathbb{R} : \underline{\mathrm{P}}(I_{\{x_E\}}[g-\mu]) \geq 0\right\} \tag{4.13}$$

for all gambles g on $\mathcal{X}_t$, **assuming that** $\overline{\mathrm{P}}(\{x_E\}) > 0$. Theorem 146172 guarantees that such inferences are coherent. The conditions that the local models should satisfy for this positivity assumption to hold are given in Proposition 145₁₇₀.

Consider the map

$$\rho_g\colon \mathbb{R}\to\mathbb{R}\colon \mu\mapsto \underline{P}(I_{\{x_E\}}[g-\mu]).$$

We know from Lemma 54_{68}, that ρ_g is continuous, concave and non-increasing. Hence $\{\mu\in\mathbb{R}\colon \rho_g(\mu)\geq 0\}=(-\infty,\underline{R}_t(g|x_E)]$, which shows that the supremum that we should have *a priori* used in (4.13) is indeed a maximum. $\underline{R}_t(g|x_E)$ is the right-most zero of ρ_g, and it is, again by separate coherence of $\underline{P}$, guaranteed to lie between the smallest value $\min g$ and the largest value $\max g$ of g. If moreover $\underline{P}(\{x_E\})>0$, then Theorem 53_{67} implies that $\underline{R}_t(g|x_E)$ is the unique zero of ρ_g. If on the other hand $\underline{P}(\{x_E\})=0$, then $(-\infty,\underline{R}_t(g|x_E)]$ is the set of all zeros of ρ_g. It appears that any algorithm for calculating $\underline{R}_t(g|x_E)$ will benefit from being able to calculate the values of ρ_g, or even more simply, to check their signs, efficiently.

4.5.2 Calculating the values of ρ_g recursively

We now recall from Section 4.3_{168} that the joint $\underline{P}$ can be constructed recursively from leaves to root. The idea we now use is that calculating $\rho_g(\mu)=\underline{P}(I_{\{x_E\}}[g-\mu])$ becomes easier if we graft the structure of the tree onto the argument $g^\mu := I_{\{x_E\}}[g-\mu]$ as follows. Define

$$g_s^\mu := \begin{cases} I_{\{x_s\}} & \text{if } s\in E\\ g-\mu & \text{if } s=t\\ 1 & \text{if } s\in\mathbb{S}\setminus(E\cup\{t\}),\end{cases}$$

then $g_s^\mu\in\mathcal{L}(\mathcal{X}_s)$ and $g^\mu=\prod_{s\in\mathbb{S}} g_s^\mu$. Also define, for any $s\in\mathbb{S}$, the gamble ϕ_s^μ on $\mathcal{X}_{\uparrow s}$ by $\phi_s^\mu := \prod_{u\in\uparrow s} g_u^\mu$. Then

$$\phi_\square^\mu = g^\mu \text{ and } \phi_s^\mu\geq 0 \text{ if } s\not\sqsubseteq t,$$

and

$$\phi_s^\mu = g_s^\mu \prod_{c\in\mathrm{ch}(s)} \phi_c^\mu \text{ for all } s\in\mathbb{S}, \tag{4.14}$$

where we use the convention that any product over an empty set of indices equals one. Equation (4.14) is the argument counterpart of Equation (4.7). Also, if $s\not\sqsubseteq t$ then g_s^μ and ϕ_s^μ do not depend on μ, nor on g. Indeed, in that case

$$\phi_s^\mu = I_{\{x_{E\cap\uparrow s}\}}. \tag{4.15}$$

First, let us consider the nodes $s \notin t$.

We define the **messages** $\underline{\pi}_s$ and $\overline{\pi}_s$ recursively by

$$\underline{\pi}_s := \underline{\mathrm{Q}}_s\Bigg(g_s^\mu \prod_{c\in\mathrm{ch}(s)} \underline{\pi}_c \Bigg| X_{\mathrm{mo}(s)}\Bigg) \text{ and } \overline{\pi}_s := \overline{\mathrm{Q}}_s\Bigg(g_s^\mu \prod_{c\in\mathrm{ch}(s)} \overline{\pi}_c \Bigg| X_{\mathrm{mo}(s)}\Bigg). \tag{4.16}$$

We summarise such a pair by the notation: $\underline{\overline{\pi}}_s := \underline{\overline{\mathrm{Q}}}_s(g_s^\mu \prod_{c\in\mathrm{ch}(s)} \underline{\overline{\pi}}_c | X_{\mathrm{mo}(s)}) := (\underline{\pi}_s, \overline{\pi}_s)$. Then there are two possibilities:

$$\underline{\pi}_s = \begin{cases} \underline{\mathrm{Q}}_s(\{x_s\}|X_{\mathrm{mo}(s)}) \prod_{c\in\mathrm{ch}(s)} \underline{\pi}_c(x_s) & \text{if } s \in E, \\ \underline{\mathrm{Q}}_s\Bigg(\prod_{c\in\mathrm{ch}(s)} \underline{\pi}_c \Bigg| X_{\mathrm{mo}(s)}\Bigg) & \text{if } s \notin E, \end{cases}$$

$$\overline{\pi}_s = \begin{cases} \overline{\mathrm{Q}}_s(\{x_s\}|X_{\mathrm{mo}(s)}) \prod_{c\in\mathrm{ch}(s)} \overline{\pi}_c(x_s) & \text{if } s \in E, \\ \overline{\mathrm{Q}}_s\Bigg(\prod_{c\in\mathrm{ch}(s)} \overline{\pi}_c \Bigg| X_{\mathrm{mo}(s)}\Bigg) & \text{if } s \notin E. \end{cases}$$

The messages $\underline{\pi}_s$ and $\overline{\pi}_s$ are gambles on $\mathcal{X}_{\mathrm{mo}(s)}$, and can therefore be seen as tuples of real numbers, with as many components $\underline{\overline{\pi}}_s(x_{\mathrm{mo}(s)})$ as there are elements $x_{\mathrm{mo}(s)}$ in $\mathcal{X}_{\mathrm{mo}(s)}$. They are all non-negative. As their notation suggests, they do not depend on the choice of g or μ, but only (at most) on which nodes are **instantiated**, i.e., belong to E, and on which value x_E the variable X_E for these instantiated nodes assumes.

It then follows from Equations (4.7) and (4.14) and the strong factorisation property[8] that

$$\underline{\mathrm{P}}_{\uparrow s}(\phi_s^\mu | X_{\mathrm{mo}(s)}) = \underline{\pi}_s \text{ and } \overline{\mathrm{P}}_{\uparrow s}(\phi_s^\mu | X_{\mathrm{mo}(s)}) = \overline{\pi}_s. \tag{4.17}$$

[8] This, together with the course of reasoning leading to Equation (4.21), shows that the results of updating the tree (and the algorithm we are deriving) in this way will be exactly the same **for any way** of forming a product of the local models for the children of s, **provided only that this product is strongly factorising**. For instance, replacing the conditionally independent natural extension with the strong product in Equation (4.6) will lead to exactly the same inferences. Of course, this should not be taken to mean that our algorithm also works for updating credal trees under strong independence.

Next, we turn to nodes $s \sqsubseteq t$.

Define the messages π_s^μ by

$$\pi_s^\mu := \underline{\mathrm{Q}}_s(\psi_s^\mu | X_{\mathrm{mo}(s)}), \tag{4.18}$$

where the gambles ψ_s^μ on $\mathcal{X}_s$ are given by the recursion relations:[9]

$$\psi_t^\mu := \max\{g-\mu,0\}\prod_{c\in\mathrm{ch}(t)}\underline{\pi}_c + \min\{g-\mu,0\}\prod_{c\in\mathrm{ch}(t)}\overline{\pi}_c, \tag{4.19}$$

and for each $\square \neq s \sqsubseteq t$, so $\mathrm{mo}\, s$ exists,

$$\psi_{\mathrm{mo}(s)}^\mu := \left[\max\{\pi_s^\mu,0\}\prod_{c\in\mathrm{sib}(s)}\underline{\pi}_c + \min\{\pi_s^\mu,0\}\prod_{c\in\mathrm{sib}(s)}\overline{\pi}_c\right] g_{\mathrm{mo}\, s}^\mu. \tag{4.20}$$

The messages π_s^μ are again tuples of real numbers, with one component $\pi_s^\mu(x_{\mathrm{mo}(s)})$ for each of the possible values $x_{\mathrm{mo}(s)}$ of $X_{\mathrm{mo}(s)}$.[10] They do depend on the choice of g or μ, as well as on which nodes are instantiated and on which value x_E the variable X_E for these instantiated nodes assumes.

It then follows from Equations (4.7) and (4.14) and the strong factorisation property of the local independent products that

$$\underline{\mathrm{P}}_{\uparrow s}(\phi_s^\mu | X_{\mathrm{mo}(s)}) = \pi_s^\mu \text{ and of course } \rho_g(\mu) = \pi_\square^\mu. \tag{4.21}$$

We conclude that we can find the value of $\rho_g(\mu)$ by a backwards recursion method consisting in passing messages up to the root of the tree, and in transforming them in each node using the local uncertainty models; see Equations (4.16) and (4.18)–(4.20).

There is a further simplification, because we are not necessarily interested in the actual value of $\rho_g(\mu)$, but rather in its sign. It arises whenever there are instantiated nodes above the target node: $E \cap \mathrm{ancest}(t) \neq \emptyset$. Let in that case e_t be the greatest element of the chain $E \cap \mathrm{ancest}(t)$, i.e., the instantiated node closest to and preceding the target node t, and let s_t be its successor in the chain $\downarrow t$; see for instance Figure 4.1$_{179}$. If we let

$$\lambda_g(\mu) := \max\{\pi_{s_t}^\mu(x_{e_t}),0\}\prod_{c\in\mathrm{sib}(s_t)}\underline{\pi}_c(x_{e_t}) + \min\{\pi_{s_t}^\mu(x_{e_t}),0\}\prod_{c\in\mathrm{sib}(s_t)}\overline{\pi}_c(x_{e_t}),$$

[9] The maximum or minimum of two gambles (the zero gamble is also a gamble) should be interpreted as a pointwise maximum, minimum respectively.

[10] If s is the root node, then $\mathrm{mo}(s) = \emptyset$ and π_s^μ is a single real number, which by Equation (4.21) is equal to $\rho_g(\mu)$. See also footnote 1$_{163}$.

then it follows from Equation (4.20) [with $s = s_t$ and $\mathrm{mo}(s) = e_t$] that $\psi^{\mu}_{e_t} = I_{\{x_{e_t}\}}\lambda_g(\mu)$. We easily derive that

$$\rho_g(\mu) = a\max\{\lambda_g(\mu), 0\} + b\min\{\lambda_g(\mu), 0\}, \tag{4.22}$$

where a and b are real constants that do not depend on g and μ. Letting $g := \mu \pm 1$ then allows us to identify the constants a and b. It is easy to see, however, that $b > 0$ and $a \geq 0$ because we assumed from the outset that $\overline{\mathrm{P}}(\{x_E\}) > 0$. We gather from this observation that

$$\underline{\mathrm{R}}_t(g|x_E) = \max\{\mu \in \mathbb{R} \colon \lambda_g(\mu) \geq 0\}.$$

Moreover, by combining Equations (4.15) and (4.17) with Proposition 145_{170}, we find that $\overline{\pi}_c(x_{e_t}) = \overline{\mathrm{P}}_{\uparrow c}(\{x_{E\cap\uparrow c}\}|x_{e_t}) > 0$ for all $c \in \mathrm{sib}(s_t)$, and therefore $\lambda_g(\mu) \geq 0 \Leftrightarrow \pi^{\mu}_{s_t}(x_{e_t}) \geq 0$. Hence

$$\underline{\mathrm{R}}_t(g|x_E) = \max\left\{\mu \in \mathbb{R} \colon \pi^{\mu}_{s_t}(x_{e_t}) \geq 0\right\}.$$

We conclude that in order to update the tree in the situation described above, we can perform all calculations on the sub-tree $\uparrow s_t$, where the new root s_t has local model $\underline{\mathrm{Q}}_{s_t}(\cdot|x_{e_t})$. This is also borne out by the discussion of the separation properties in Section 4.4_{173}.

4.5.3 The algorithm

We now convert these observations into a workable algorithm.

Using regular extension and message passing, we are able to compute $\underline{\mathrm{R}}_t(g|x_E)$: we (i) choose any $\mu \in [\min g, \max g]$; (ii) calculate the value of $\lambda_g(\mu)$ by sending messages from the terminal nodes towards the root; and (iii) repeat this in some clever way to find the maximal μ that will make this $\lambda_g(\mu)$ zero. But we have seen above that this naive approach can be sped up by exploiting (a) the separation properties of the tree, and (b) the independence of μ (and g) for some of the messages, namely those associated with nodes that do not precede the target node t.

For a start, as we are only interested in the sign of $\rho_g(\mu)$ [or equivalently, that of $\lambda_g(\mu)$], which we have seen is determined by the sign of $\pi^{\mu}_{s_t}(x_{e_t})$, we only have to take into consideration nodes that strictly follow e_t.

The next thing a smarter implementation of the algorithm can do, is determine the **trunk** $\tilde{\mathbb{S}}$ of the tree: those nodes that precede the queried

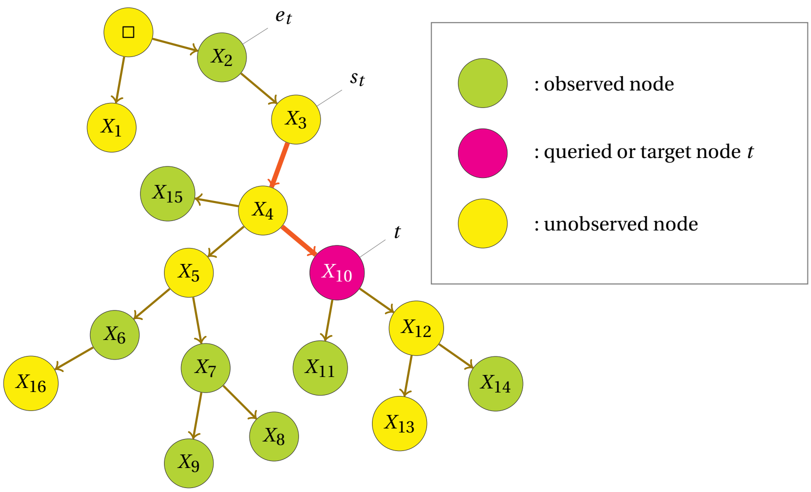


Figure 4.1: Example imprecise Markov tree. The target node is $t = 10$, $e_t = 2$ is the 'greatest' observed ancestor of t and $s_t = 3$ is the child of e_t that precedes t. The bolder arrows represent the trunk $\tilde{\mathbb{S}} = \{3, 4, 10\}$ of the tree.

node t and strictly follow the greatest observed node e_t preceding t. We can define the trunk more formally as follows: $\tilde{\mathbb{S}} := \downarrow t \cap \uparrow \mathrm{ch}(e_t)$. For the tree in Figure 4.1 for instance, where the darker X_{10} is the queried variable and the lighter nodes $\{2, 6, 7, 8, 9, 11, 14, 15\}$ are instantiated, the trunk is given by $\tilde{\mathbb{S}} = \{3, 4, 10\}$, and indicated by bolder arrows.

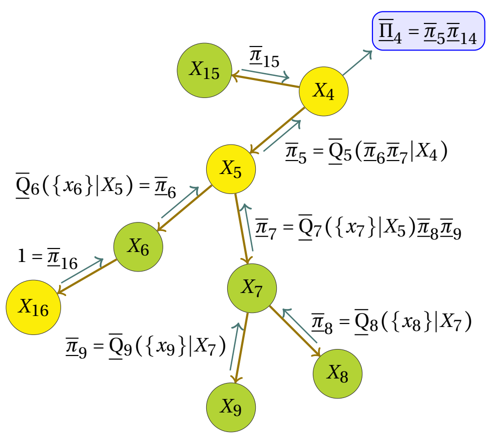


Figure 4.2: Calculation of $\overline{\Pi}_4$, which is a summary of the μ-independent messages in the trunk node 4.

We have a special interest in the nodes that constitute the trunk, because only they will send messages to their mother nodes that actually depend on μ. As a consequence, all other nodes (all descendants of the trunk that are not in the trunk themselves) send messages that have to be calculated only once. This implies that we can summarise all the μ-independent messages by propagating all of them until they reach the trunk. The μ-independent messages $\underline{\pi}_s$ that arrive in a trunk node s can be represented more succinctly by their point-wise products $\underline{\Pi}_s \coloneqq \prod_{c\in\mathrm{ch}(s)\setminus\tilde{\mathbb{S}}} \underline{\pi}_c$, because Equations (4.19) and (4.20) only depend on them through these products.

This means that for every trunk node $s \in \tilde{\mathbb{S}}$, we have to find the lower (upper) messages of every child c of s that is not in the trunk itself. Both $\underline{\pi}_c$ and $\overline{\pi}_c$ can be calculated recursively using Equation (4.17), where the recursion starts at the leaves and moves up to (but stops right before) the trunk. In the leaves, the local lower and upper previsions of the indicator of the evidence are sent upwards if the leaf is instantiated; if not the constant 1 is sent up, which is equivalent to deleting the node from the tree. We could envisage removing **barren nodes** (all of whose descendants are uninstantiated, such as X_1, X_{13}, X_{16} in the example tree above) from the tree beforehand, but we believe the computational overhead created by the search for them will void the gain.

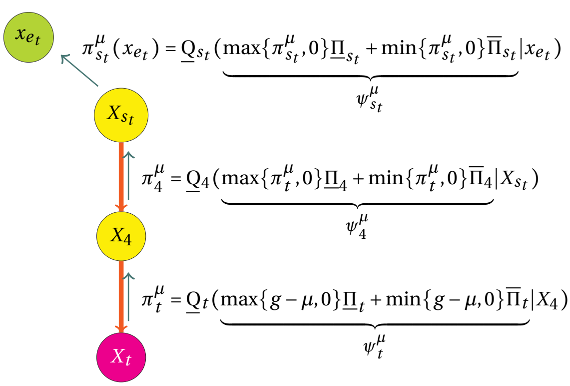


Figure 4.3: Calculation of $\pi^\mu_{s_t}(x_{e_t})$, whose sign is the same as that of the lower prevision $\underline{P}(I_{\{x_E\}}[g-\mu])$.

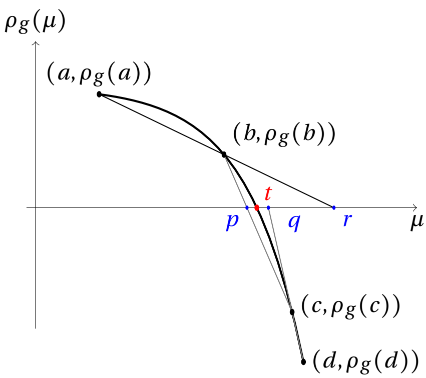
The only recursion that is still left to do, is the calculation of the μ-dependent messages π^μ_s along the trunk. As demonstrated in Figure 4.3, we can calculate $\pi^{\mu}_{s_t}(e_t)$ using the following recursion formula:

$$\pi^{\mu}_{s} := \begin{cases} \underline{\mathrm{Q}}_s(\max\{g-\mu,0\}\underline{\Pi}_s + \min\{g-\mu,0\}\overline{\Pi}_s | X_{\mathrm{mo}(s)}) & s = t, \\ \underline{\mathrm{Q}}_s(\max\{\pi^{\mu}_{c_t},0\}\underline{\Pi}_s + \min\{\pi^{\mu}_{c_t},0\}\overline{\Pi}_s | X_{\mathrm{mo}(s)}) & s \in \tilde{\mathbb{S}} \setminus \{t\} \text{ and } \\ & \mathrm{ch}(s) \cap \tilde{\mathbb{S}} = \{c_t\}. \end{cases}$$

These formulas are reformulations of Equations (4.18)–(4.20), where the influence of the $\overline{\underline{\Pi}}$ has been made explicit.

Since we now know how to calculate $\pi^{\mu}_{s_t}(e_t)$, we can tackle the final problem: find the maximal μ for which $\pi^{\mu}_{s_t}(e_t) = 0$. In principle, a secant root-finding method could be used, but using the concavity and non-increasing character of $\pi^{\mu}_{s_t}(e_t)$ as a function of μ, we can speed up the calculation of the maximal root drastically as shown in Figure 4.4.

Let us briefly discuss the complexity of our algorithm. Consider for a start that for a fixed μ each node makes a single local computation and then propagates the result to its mother node: this implies that, with μ fixed, the algorithm is linear in the number of nodes. Iterating on μ then amounts to multiplying such a linear complexity with the number of iterations. This



$p := \frac{\rho_g(c)b - \rho_g(b)c}{\rho_g(c) - \rho_g(b)}$;
$m := c$; $t := \frac{p+m}{2}$;
while $m - p > tol$ and $\rho_g(t) \neq 0$
if $\rho_g(t) > 0$
$a := b$; $b := t$;
$s := \frac{\rho_g(a)b - \rho_g(b)a}{\rho_g(a) - \rho_g(b)}$;
else
$d := c$; $c := t$;
$s := \frac{\rho_g(c)d - \rho_g(d)c}{\rho_g(c) - \rho_g(d)}$;
$p := \frac{\rho_g(c)b - \rho_g(b)c}{\rho_g(c) - \rho_g(b)}$;
$m := \min\{m, s\}$; $t := \frac{p+m}{2}$;

Figure 4.4: The root of a concave and non-increasing function ρ_g whose values $\rho_g(a) > \rho_g(b) > 0 > \rho_g(c) > \rho_g(d)$ are known, will always be in the interval $[p, m]$ with $m := \min\{q, r\}$. Here p, q and r are the intersections with the horizontal axis of the straight lines through $(b, \rho_g(b))$ and $(c, \rho_g(c))$, $(c, \rho_g(c))$ and $(d, \rho_g(d))$, and $(a, \rho_g(a))$ and $(b, \rho_g(b))$, respectively. The next function evaluation of ρ_g will be in t which bisects the error interval $[p, m]$. If $\rho_g(t) > 0$, then a becomes b and b becomes t, otherwise d becomes c and c becomes t and a new interval $[p, m]$ and matching t can be calculated. We stop iterating as soon as the error interval $[p, m]$ is smaller than a given tolerance *tol*, or $\rho_g(t)$ is exactly zero.

number depends on the function g, as the iterations are made to compute the root of a function that is known to belong to the real interval $[\min g, \max g]$. If we assume that the bisection algorithm is employed to find the root—for the sake of simplicity—and let $r := \max g - \min g$ be the range of the function, then the number of iterations is bounded by $\log_2 \frac{r}{tol} + 1$, where tol is some fixed tolerance. In other words, the number of iterations is linear in the number b of bits needed to represent r. This means that the overall complexity of the algorithm is $O(b \cdot |\mathbb{S}|)$, taking into account that the computational complexity of our root-finding algorithm must be lower than for the bisection (and actually also for the secant) algorithm. Since b will be a small number in most cases (e.g. when the focus is on probabilities), we simply refer to the complexity of our algorithm as linear in the number of nodes.

4.6 A simple example involving dilation

We present a very simple example that allows us to (i) follow the inference method discussed above in a step-by-step fashion; (ii) see that there are separation properties for credal nets under strong independence that fail for credal trees under epistemic irrelevance; and (iii) see that in that case we will typically observe dilation.

Consider the following imprecise Markov chain:

$$\begin{tikzcd} X_1 \arrow[r] \arrow[d, dotted] & X_2 \arrow[r] & X_3 \\ ? & x_2 \arrow[u, dotted] & x_3 \arrow[u, dotted] \end{tikzcd}$$

To make things as simple as possible, we suppose that $\mathcal{X}_1 = \{a, b\}$ and that $\underline{Q}_1$ is a linear (or precise, or expectation-like) model Q_1 with mass function q. We also assume that $\underline{Q}_2(\cdot|X_1)$ is a linear model $Q_2(\cdot|X_1)$ with conditional mass function $q(\cdot|X_1)$. We make no such restrictions on the local model $\underline{Q}_3(\cdot|X_2)$. We also use the following simplifying notational device: if we have three real numbers $\underline{\kappa}$, $\overline{\kappa}$ and γ, we let

$$\underline{\overline{\kappa}}\langle\gamma\rangle := \underline{\kappa} \max\{\gamma, 0\} + \overline{\kappa} \min\{\gamma, 0\}.$$

We observe $X_2 = x_2$ and $X_3 = x_3$, and want to make inferences about the target variable X_1: for any $g \in \mathcal{L}(\mathcal{X}_1)$, we want to know $\underline{R}_1(g|x_{\{2,3\}})$. Letting $\underline{r} :=$

$\underline{\mathrm{R}}_1(\{a\}|x_{\{2,3\}})$ and $\overline{r} := \overline{\mathrm{R}}_1(\{a\}|x_{\{2,3\}})$, we infer from the separate coherence of $\underline{\mathrm{R}}_1(\cdot|x_{\{2,3\}})$ that it suffices to calculate $\underline{r}$ and $\overline{r}$, because

$$\underline{\mathrm{R}}_1(g|x_{\{2,3\}}) = g(b) + \overline{\underline{r}}\langle g(a) - g(b)\rangle.$$

We let $g^\mu = [I_{\{a\}} - \mu] I_{\{x_2\}} I_{\{x_3\}}$, and apply the approach of the previous section. We see that the trunk $\tilde{\mathbb{S}} = \{1\}$, and the instantiated leaf node 3 sends up the messages $\overline{\underline{\pi}}_3 = \overline{\underline{\mathrm{Q}}}_3(\{x_3\}|X_2)$ to the instantiated node 2, which transforms them into the messages

$$\overline{\underline{\pi}}_2 = \overline{\underline{\mathrm{Q}}}_2(\{x_2\}|X_1)\overline{\underline{\pi}}_3(x_2) =: q(x_2|X_1)\overline{\underline{q}},$$

where we let $q(x_2|X_1) := \overline{\underline{\mathrm{Q}}}_2(\{x_2\}|X_1)$ and $\overline{\underline{q}} := \overline{\underline{\pi}}_3(x_2)$. These messages are sent up to the (target) root node $t = 1$, which transforms them into the message $\pi_1^\mu = \mathrm{Q}_1(\psi_1^\mu)$ with $\psi_1^\mu = q(x_2|X_1)\overline{\underline{q}}\langle I_{\{a\}} - \mu\rangle$. If we also use that $0 \le \mu \le 1$, this leads to

$$\underline{\mathrm{P}}_1(g^\mu) = \pi_1^\mu = q(a)q(x_2|a)\underline{q}[1-\mu] + q(b)q(x_2|b)\overline{q}[-\mu],$$

so we find after applying regular extension that

$$\begin{aligned}
\underline{r} = \underline{\mathrm{R}}_1(\{a\}|x_{\{2,3\}}) &= \frac{q(a)q(x_2|a)\underline{q}}{q(a)q(x_2|a)\underline{q} + q(b)q(x_2|b)\overline{q}}\\
\overline{r} = \overline{\mathrm{R}}_1(\{a\}|x_{\{2,3\}}) &= \frac{q(a)q(x_2|a)\overline{q}}{q(a)q(x_2|a)\overline{q} + q(b)q(x_2|b)\underline{q}}.
\end{aligned}$$

When $\underline{q} = \overline{q}$, which happens for instance if the local model for X_3 is precise, then we see that, with obvious notations,

$$\overline{r} = \underline{r} = \frac{q(a)q(x_2|a)}{q(a)q(x_2|a) + q(b)q(x_2|b)} =: p(a|x_2) \tag{4.23}$$

and therefore X_2 indeed separates X_3 from X_1. But in general, letting $\alpha := q(a)q(x_2|a)$ and $\beta := q(b)q(x_2|b)$, we get

$$\overline{r} - p(a|x_2) = \frac{\alpha\beta}{\alpha+\beta}\frac{\overline{q}-\underline{q}}{\alpha\overline{q}+\beta\underline{q}} \text{ and } p(a|x_2) - \underline{r} = \frac{\alpha\beta}{\alpha+\beta}\frac{\overline{q}-\underline{q}}{\alpha\underline{q}+\beta\overline{q}}.$$

As soon as $\overline{q} > \underline{q}$, X_2 no longer separates X_3 from X_1, and we witness **dilation** [30, 50]—the increase of uncertainty on extra observations, no matter what they are—because of the additional observation of X_3!

CHAPTER 5

Conclusions

"Chapeau" and I thank you very much if you read yourself all the way through this thesis. It is OK if you did not, I am still very happy that you are about to read this conclusion. The hat might stay though. I wish you an entertaining stroll through these conclusions which is a small summary with some additional remarks about the modelling uncertainty chapter and the chapter probability trees and the chapters: Markov chains and Markov trees with the, admittedly, dull and uninspired predicate "imprecise".

Modelling uncertainty

In Chapter 1_{28}, we introduced sets of acceptable gambles in an attempt to define an uncertainty model that allows for indecision. The agent is offered a number of gambles and he can either accept them or not accept them. We assumed that the underlying utility is linear and that the agent meets a minimal number of rationality requirements, like accepting sure gain and avoiding Dutch book scenarios. A set of acceptable gambles that does not show irrational behaviour is called a coherent set of gambles. Our approach is strongly influenced by de Finetti [24], Ramsey [48], Williams [69] and Walley [62, 64]. Besides the mathematical simplicity, we prefer to work with sets of gambles because of their clear interpretation and operational nature.

In Section 1.3_{42} we developed a multivariate framework that shows how sets of acceptable gambles (modelling beliefs of the same agent) should be combined. When we want to know what exactly are the agent's beliefs about a subset of the modelled random variables, then it is explained in Section 1.3.1[43] how beliefs can be marginalised and Section 1.3.2[45] explains how to update beliefs on events. Apart from some pathological cases, both marginalising and updating result in intersecting the cone of acceptable gambles with a suitable subspace. This also explains why marginalising and updating are commutative operations.

Coherent sets of fully resolved gambles can be seen as the acceptability counterpart of linear previsions. We prove an acceptability version of Walley's lower envelope theorem [62, §2.6.3]. The set of all dominating resolved coherent sets of acceptable gambles appearing in this lower envelope theorem, can be seen as the acceptability analogue of credal sets. Usually it is assumed that a credal set is closed. This closedness can be achieved by restricting ourselves to sets of marginally acceptable gambles. Sets of marginally acceptable gambles are one of the types of uncertainty models we derived from sets of acceptable gambles. The other ones are sets of strictly desirable gambles and lower (or upper) previsions. The three derived models lack some of the expressiveness that comes with sets of acceptable gambles. Therefore there is no bijective relation between lower previsions and sets of acceptable gambles. In general this is not a problem. Only when updating, this might cause problems and this is exactly where the sets of marginally desirable and sets of strictly desirable gambles come in as they are special extreme cases.

By asking an agent about his disposition towards a gamble f and its negation $-f$, we are able to make the distinction between desirability and irrelevance. By doing so, we show that we developed a theory that deals with strong symmetry in a very natural way. Symmetry was not the focus of this thesis, and this result is merely a by-product. However, we believe that this elegant way of describing symmetry might be the strongest point in favour of the framework of sets of acceptable gambles in future work.

Although finite possibility spaces are assumed, we do not expect major difficulties when extending our theory to infinite possibility spaces. The avoiding sure loss axiom will have to be adapted slightly (see Chapter 2[80]), and the Axiom of Choice will have to be used a number of times; sadly this means that the constructive nature of some of the proofs will be lost. Unfortunately, the results about the combination of models will not be that easy to transfer when considering infinite possibility spaces. For example, it is known that (cut) conglomerability has to be assumed for marginal extension.

Imprecise probability trees

In Chapter 2_{80}, we have generalised probability trees—arguably the simplest existing graphical model—in a way that is strongly connected to a special case of Shafer and Vovk's game-theoretic probability.

We haven't paid much attention to the special case that the coherent lower previsions and their conjugate upper previsions coincide, and are therefore (precise) **previsions** or **fair prices** in de Finetti's [24] sense. When all the local predictive models $\underline{\mathrm{P}}_t$ (see Proposition 97_{107}) happen to be precise, meaning that $\underline{\mathrm{P}}_t(f) = \overline{\mathrm{P}}_t(f) = -\underline{\mathrm{P}}_t(-f)$ for all gambles f on $\mathrm{ch}(t)$, then the immediate prediction model we have described becomes very closely related, and arguably identical to, the probability trees introduced and studied by Shafer in [53]. Indeed, we then get predictive previsions $\mathrm{P}(\cdot|s)$ that can be obtained through concatenation of the local models P_t, as guaranteed by Theorem 96_{106}.

We want to recall that Theorem 96_{106} and Proposition 97_{107} allow for a calculation of the predictive models $\underline{\mathrm{P}}(\cdot|s)$ using only the local models and **backwards recursion**, in a manner that is strongly reminiscent of dynamic programming techniques. This should allow for a much more efficient computation of such predictive models than, say, an approach that exploits lower envelope theorems and sets of probabilities/previsions and forms the basis for efficient algorithms when dealing with other types of graphical models.

What makes this more efficient approach possible is, ultimately, the Marginal Extension Theorem (Theorem 30_{51}), which leads to the Concatenation Formula (Theorem 96_{106}). Generally speaking (see for instance [62, § 6.7] and [41]), such marginal extension results can be proved because of the hierarchical and local nature of the assessments.

In Chapter 2_{80}, we also give some examples, like the Gambler's ruin which is a special instantiation of a counting process. General counting processes are something that cannot be handled in our theory because we do not allow for gambles that become infinite. Another example boils down to the irrelevant forward product, which corresponds to the epistemically independent product when permutation-invariant gambles are considered.

This independent product is definitely something that deserves a bit more of study as it forms the basis of every statistical theory.

In the last part of this chapter, sub- and supermartingales are considered. Submartingales form a basic tool and concept in Shafer and Vovk's game-theoretic probability theory, There is a strong connection [12] between the theory of imprecise event trees and Shafer and Vovk's game-theoretic probability. In the case of finite depth trees, the connection was proved [12], but it remains still open whether the Matching Theorem in [12] holds in general, for unbounded depth trees.

Imprecise Markov chains

In Chapter 3_{122}, a special type of imprecise probability tree was studied: the imprecise Markov chain. It can be viewed as a generalisation of classical discrete time Markov chains with finite state space in which the Markov condition is interpreted in terms of Walley's epistemic irrelevance. This interpretation does not entirely coincide with the sensitivity analysis interpretation that was used in previous attempts to describe Markov chains with parameter uncertainty. In the sensitivity analysis interpretation, the Markov condition is interpreted as a strong independence assessment and transition probabilities are replaced with credal sets, so these types of generalised Markov chains are effectively special types of credal networks. Nevertheless, both approaches yield the same results if we restrict ourselves to calculating the marginal upper expectations for variables X_n. But in any case, for the actual calculation of expectations, the set of transition matrices approach suffers from a combinatorial explosion and the resulting high computational complexity. This can be avoided using our upper transition operator approach. We have proved that under the epistemic irrelevance Markov assumption, lower or upper previsions of gambles can be calculated in a recursive fashion. When calculating marginal gambles, the complexity of this approach even becomes linear when the transition operator is considered an oracle.

An important situation where the interpretation of the Markov condition is not important is in the study of the limit behaviour—and closely linked to that—ergodicity of upper transition operators. We have given different equivalent conditions under which an upper transition operator is ergodic. We have shown that ergodicity is completely determined by the eigenvalues and functions of the transition operator as is the case in classical Markov chains. Unfortunately, it is at this point not known how to calculate these eigenvalues in general. This is why we developed an alternative test for ergodicity, which needs at most $2|\mathcal{X}| - 1$ evaluations of the upper transition operator. Any algorithm that implements this test consists of two steps: the first checks for top class regularity by building the upper accessibility graph and checking for final strongly connected components and their cyclicity. In some cases a second step is needed, to check for top class absorption.

Another approach that has been documented in the literature [61], calculates the coefficient of ergodicity and checks whether there is some iterate of the transition operator such that the corresponding coefficient becomes strictly smaller than one. If this is the case, then the non-expansive map that every upper transition operator is, becomes a contractive map and ergodicity is a fact. Interesting about the coefficient of ergodicity is that it moreover provides an upper bound on the speed of convergence. What makes this approach difficult to use outside a theoretical context, is that there is at present no efficient algorithm to calculate the coefficient of ergodicity. It is moreover likely that very high powers of the upper transition operator need to be calculated.

A paper with a different background is the very general work of Akian and Gauber [1], who describe an algorithm for checking ergodicity of upper transition operators that are piecewise affine. In practice, their algorithm relies heavily on extreme points to calculate the subdifferential. If the set of extreme points is given, then their critical graph approach is the shortest way to get to all qualitative information available on the eigenspace of the upper transition operator. If these extreme points are not given explicitly, then a vertex enumeration step is involved which is computationally very hard as any algorithm based on vertex enumeration cannot have polynomial time complexity.

Our algorithm avoids the vertex enumeration step by using the upper transition operator directly. It also allows checking for ergodicity for upper transition operators whose 'credal set' has an infinite number of extreme points. Of course, extra information about the eigenspace available through the critical graph approach, not necessary for deciding upon ergodicity, may be lost by using our simpler approach based on accessibility alone.

In a number of stochastic control applications that provide a motivation for Akian and Gaubert's work [1], the extreme points of the polytopes of transition probability measures cannot be enumerated (only separation or minimisation oracles are available), and hence, dealing with such situations in the way we explain here, is also quite relevant in that application context.

Imprecise Markov trees

In Chapter 4_{160}, we have defined imprecise-probability (or credal) trees using Walley's notion of epistemic irrelevance. Credal trees generalise tree-shaped Bayesian nets in two ways: by allowing the parameters of the tree to be imprecisely specified, and moreover by replacing the notion of stochastic independence with that of epistemic irrelevance. Our focussing on epistemic irrelevance is the most original aspect of this work, as this notion has received limited attention so far in the context of credal nets.

We have focussed in particular on developing an efficient exact algorithm for updating beliefs on the tree. Like the algorithms developed for precise graphical models, our algorithm works in a distributed fashion by passing messages along the tree. It computes lower and upper conditional previsions (expectations) with a complexity that is essentially linear in the number of nodes in the tree. This is remarkable because until now it was unclear whether an algorithm with such features was at all feasible: in fact, epistemic irrelevance is most easily formulated using coherent lower previsions or sets of acceptable gambles, which have never before been used as such in practical applications of credal nets, which tend to focus on a sets of probabilities approach. Moreover, it is at this point not clear that epistemic irrelevance is as "well-behaved" as strong independence is with respect to the graphoid axioms for propagation of probability in graphical models [10, 42].[1] Our results therefore appear very encouraging, and seem to have the potential to open up new avenues of research in credal nets.

On a more theoretical side, we have also shown that our credal trees satisfy the important rationality requirement of coherence. This has been established under the assumption that the **upper** probability of any possible observation in the tree is positive, which is a very mild requirement. The same assumption also allowed us to show that all inferences made by updating

[1] Unlike credal nets based on strong independence, a credal net based on epistemic irrelevance cannot generally be seen as equivalent with a set of Bayesian nets **with the same graphical structure**: if it were, then all separation properties of Bayesian nets would simply be inherited, and we have seen in Section 4.6_{182} that such is not the case.

the tree will be coherent with each other as well as with the local uncertainty models in the nodes of the tree.

Where to go from here? There are many possible avenues for future research.

It would be very useful to be able to extend the algorithm at least to so-called **polytrees**, which are substantially more expressive graphs than trees are. This could be a difficult task to achieve. In fact, updating credal nets based on strong independence is an NP-hard task when the graph is more general than a tree [11]. Similar problems might affect the algorithms for credal nets based on irrelevance.

For applications, it would be very important to develop statistical methods specialised for credal nets under irrelevance that avoid introducing excessive imprecision in the process of inferring probabilities from data. This could be achieved, for instance, by using a single global IDM [63] over the variables of the tree rather than many local ones.

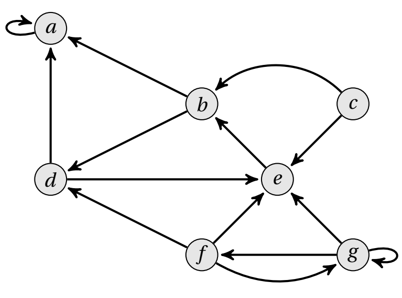
Another research direction could be concerned with trying to strengthen the conclusions that epistemic trees lead to. There might be cases where our Markov condition based on epistemic irrelevance is too weak as a structural assessment. We have discussed situations where this type of Markov condition systematically leads to a dilation of uncertainty when updating beliefs with observations, and indicated that this dilation is related to (the lack of) certain separation properties induced by epistemic irrelevance on a graph. Dilation might not be desirable in some applications, and we could be called upon to strengthen the model in order to rule out such behaviour. One way to address the issue of dilation—but not necessarily the easiest—could consist in adding additional irrelevance statements to the model, other than those derived from the Markov condition. An easier avenue could be based on designing assumptions that together with the Markov condition lead to some stronger separation properties, while not necessarily requiring them to match the common ones used in Bayesian nets.

APPENDIX **A**

The very basics of Graph Theory

In this appendix we will introduce the graph-theoretical notation that is used in this work. We base ourselves on the work of Bang-Jensen & Gutin [3] and Jarvis & Shier [33].

A directed graph $\mathcal{G}$ consists of a non-empty finite set $V(\mathcal{G})$ of elements called **vertices** or sometimes called nodes, and a finite set $A(\mathcal{G})$ of ordered pairs of vertices called **arcs**. In the figure we have a directed graph with set of vertices $V(\mathcal{G}) = \{a,b,c,d,e,f,g\}$ and set of arcs $A(\mathcal{G}) = \{(a,a),(b,a),(c,b),(d,a),(b,d),(e,b),(c,e),(d,e),(f,d),(f,e),(g,e),$ $(g,f),(f,g),(g,g)\}$.

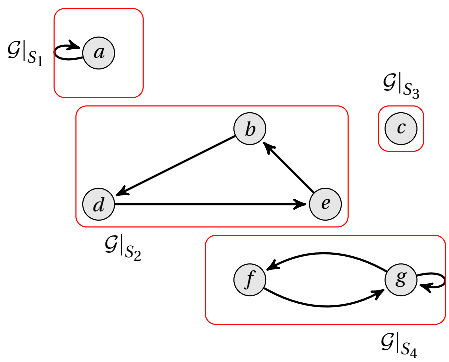


Given an arc (x,y) of a graph $\mathcal{G}$ we say that the vertex x is the head and y is the tail of the arc. A **path** of length k in a graph $\mathcal{G}$ is a sequence of vertices $x_1x_2\ldots x_{k-1}x_k$ such that $(x_i,x_{i+1}) \in A(\mathcal{G})$ for all $i \in \{1,2,\ldots,k-1\}$ and such that no vertex in the sequence $x_1x_2\ldots x_{k-1}$ gets repeated. Any path in $\mathcal{G}$ of length k with $k \in \mathbb{N}_{>0}$ from x to y is simply called a path in $\mathcal{G}$ and is denoted an (x,y)-path. An (x,x)-path is also called a **cycle**. In the example above, eba and $fgeb$ are paths while $fgfeb$ and $fgge$ are not; fgf, aa and $debd$ are cycles.

A **subgraph** $\mathcal{H}$ of $\mathcal{G}$ is a directed graph with $V(\mathcal{H}) \subseteq V(\mathcal{G})$ and $A(\mathcal{H}) \subseteq A(\mathcal{G})$. We say that the subgraph $\mathcal{H}$ of $\mathcal{G}$ is **induced by** $S = V(\mathcal{H})$ if $A(\mathcal{H}) = A(\mathcal{G}) \cap V(\mathcal{H})^2$. We denote the subgraph of $\mathcal{G}$ induced by S as $\mathcal{G}|_S$.

A.0.1 Strongly connected components

A vertex y in a graph $\mathcal{G}$ is **reachable** from x if $x = y$ or if there is an (x,y)-path in $\mathcal{G}$. The reachability relation on $V(\mathcal{G})$ is a preorder, i.e., it is reflexive and transitive. The associated equivalence relation "x is reachable from y and y is reachable from x" partitions the vertices $V(\mathcal{G})$ into equivalence classes $S_1, S_2, \ldots, S_n$ and the induced subgraphs $\mathcal{G}|_{S_i}$ are the **strongly connected components** or **strong components** of $\mathcal{G}$. Every vertex in a strongly connected component is reachable by any other vertex. If there is only one strongly connected component, then we say that the graph is strongly connected or strong. In this case, every vertex is reachable from every other vertex. In the example graph, the strongly connected components are the subgraphs induced by $S_1 = \{a\}$, $S_2 = \{b,d,e\}$, $S_3 = \{c\}$ and $S_4 = \{f,g\}$.



Tarjan [57] has developed an algorithm that is widely used to compute strongly connected components of a graph $\mathcal{G}$ and that has time complexity $\mathcal{O}(|V(\mathcal{G})| + |A(\mathcal{G})|)$.

The **period** of a strongly connected network is the greatest common divisor of the lengths of the cycles it contains.

APPENDIX **B**

A short introduction to Order Theory

We will try to give a short overview of the order theoretic concepts used in this thesis.

B.1 Partial orders and their representation

Definition 147: *A set P with a binary relation $\leq$ is **partially ordered** if it is reflexive, antisymmetric and transitive, i.e. if for all $x, y, z \in P$*

PO1. $x \leq x$ — $\leq$ is **reflexive** on P;

PO2. $x \leq y$ *and* $y \leq x$ *implies* $x = y$ — $\leq$ is **antisymmetric** on P;

PO3. $x \leq y$ *and* $y \leq z$ *implies* $x \leq z$ — $\leq$ is **transitive** on P.

A partially ordered set, or shortly **poset**, will be denoted as $(P, \leq)$. When either $x \leq y$ or $y \leq x$ then we say that the pair (x, y) is **comparable**. If every possible pair of elements of a partially ordered set $(P, \leq)$ is comparable then we say that the set P is **totally ordered** by $\leq$.

A **strict order** relation $<$ can be defined from the order relation $\leq$ by demanding that $x < y$ if and only if $x \leq y$ and $x \neq y$. Both $\leq$ and $<$ can be used as fundaments for order theory. A new poset $(P, \geq)$ is induced from $(P, \leq)$ by defining $x \geq y \Leftrightarrow y \leq x$ for all $x, y \in P$.

A **Hasse-diagram** is a graphical representation of a finite partial order $(P, \leq)$. The graph is a directed graph, where the nodes are the elements of P and a directed arc is drawn from the node x to the node y, if $x < y$ and there is no element $z \in P$ such that $x < z < y$. So, a Hasse-diagram can be seen as a minimal graphical representation of a finite poset. Usually, the Hasse-diagram is defined as an undirected graph, where the direction of the order relation can be understood from the fact that greater elements are put higher in the graph. We will not follow this custom and define the Hasse-diagram immediately as a directed graph.

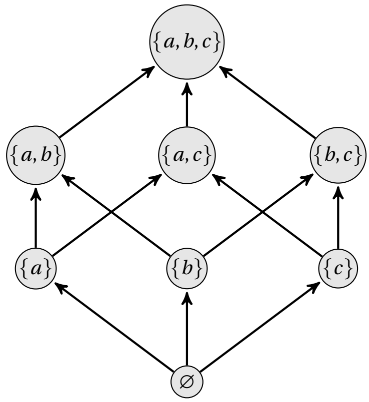


Figure B.1: The Hasse diagram of the partial order $(2^{\{a,b,c\}}, \subseteq)$.

B.1.1 Chains and antichains

Definition 148: *A **chain** C of a partially ordered set $(P, \leq)$ is a totally ordered subset of P that is not empty:*

$$C \text{ is a chain} \Leftrightarrow (\forall x, y \in C \neq \varnothing)(x \leq y \text{ or } y \leq x).$$

Definition 149: *An **antichain** C of a partially ordered set $(P, \leq)$ is a subset of P that is pairwise **incomparable**:*

$$C \text{ is an antichain} \Leftrightarrow (\forall x, y \in C)(x \nleq y \text{ and } y \nleq x).$$

An example of a chain in Figure B.1 is the set $\{\varnothing, \{b\}, \{a, b, c\}\}$. An example of an antichain in Figure B.1 is the set $\{\{a, b\}, \{a, c\}, \{b, c\}\}$.

B.2 Special elements in order relations

A (subset of a) poset $(P, \leq)$ has a **greatest element** or **maximum** g, if for every element $t \in P$ it holds that $t \leq g$. The greatest element can be considered as an upper bound of the set and is necessarily unique but does not need to exist. An upper bound of a poset $(P, \leq)$ can be defined as any set S such that for any $s \in S$ and for any $p \in P$ it holds that $p \leq s$. The maximum of a set should not be confused with the **maximal elements** of a partial order which are all the elements of the set that are undominated: $x \in P$ is maximal if and only if $z \geq x$ for any $z \in P$ implies that $z = x$.

The **supremum** or **least upper bound** of a subset S denoted $\sup S$ of a partial order $(P, \leq)$ is the smallest element of P that is greater or equal than every element of S. The supremum does not need to exist. If it does and if it is in S then it is necessarily the greatest element of S.

Given a poset $(P, \leq)$, the dual relation $\geq$ can always be defined as $x \leq y \Leftrightarrow y \geq x$. It turns out that $(P, \geq)$ is again a partial order. Similar to the definitions above **smallest element**, **minimal elements** and **minimum**, **lower bound** and **infimum** can be defined.

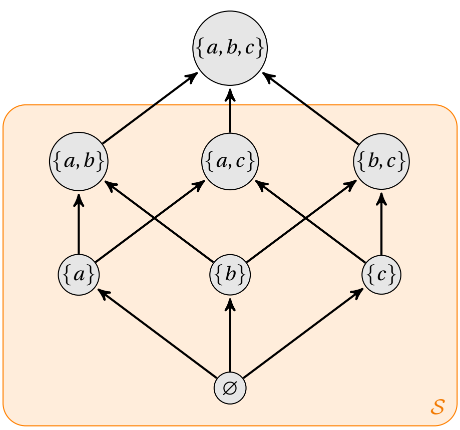


Figure B.2: The Hasse diagram of the partial order $(2^{\{a,b,c\}}, \subseteq)$. The set $\mathcal{S}$ does not have a greatest element or maximum. It does have a least upper bound though: $\sup \mathcal{S} = \{a, b, c\}$ which is the maximum of $2^{\{a,b,c\}}$. The smallest element of $\mathcal{S}$ is given by $\min \mathcal{S} = \varnothing$. Clearly, $\inf \mathcal{S} = \min \mathcal{S}$. The maximal elements of the partial order $(\mathcal{S}, \subseteq)$ are given by $\{\{a, b\}, \{a, c\}, \{b, c\}\}$.

B.3 Some special orders

We already saw one special type of partial order, the total order. Other special types of orders exist as well and we will here define some that are used in this work.

Definition 150: *A totally ordered set $(P, \leq)$ is **well-ordered** if and only if every nonempty subset $\mathcal{S}$ of P has a least element in this ordering.*

A well-ordered set is sometimes said to be a **well-founded order**. An example of a well-ordered set is the set of natural numbers $\mathbb{N}_{\geq 0}$. The set of all non-negative real numbers $\mathbb{R}_{\geq 0}$ on the other hand, is an example of a totally ordered set that is not well-ordered.

Definition 151:

1. *A set P with order relation $\leq$ is a **lattice** if every two-element subset of P has a supremum and an infimum in P.*

2. *A poset $(P, \leq)$ is a **complete lattice** if every non-empty subset has a supremum and an infimum in P.*

The set of all positive real numbers $\mathbb{R}_{>0}$, is an example of a complete lattice that is not well-ordered.

Bibliography

25, 26, 123, 124, 132, 154, 188, 189 [1] M. Akian & S. Gaubert. ‘Spectral theorem for convex monotone homogeneous maps, and ergodic control’. *Nonlinear Analysis-theory Methods & Applications* 52.2 (2003), pp. 637–679. DOI: `10.1016/S0362-546X(02)00170-0`.

53 [2] R. Bagnara, P. M. Hill & E. Zaffanella. ‘The Parma Polyhedra Library: Toward a complete set of numerical abstractions for the analysis and verification of hardware and software systems’. *Science of Computer Programming* 72.1-2 (2008), pp. 3–21. DOI: `10.1016/j.scico.2007.08.001`.

192 [3] J. Bang-Jensen & G. Gutin. *Digraphs: Theory, Algorithms and Applications.* London: Springer-Verlag, 2000.

24 [4] G. Boole. *The Laws of Thought.* New York: Dover Publications, 1847, reprint 1961.

160 [5] C. P. de Campos & F. G. Cozman. ‘Computing lower and upper expectations under epistemic independence’. *International Journal of Approximate Reasoning* 44.3 (2007), pp. 244 –260. DOI: `10.1016/j.ijar.2006.07.013`.

40 [6] I. Couso & S. Moral. ‘Sets of desirable gambles: Conditioning, representation, and precise probabilities’. *International Journal of Approximate Reasoning* 52.7 (2011), pp. 1034–1055. DOI: `10.1016/j.ijar.2011.04.004`.

[7] I. Couso, S. Moral & P. Walley. 'A survey of concepts of independence for imprecise probabilities'. *Risk, Decision and Policy* 5.2 (2000), pp. 165–181. 174

[8] F. G. Cozman. 'Credal networks'. *Artificial Intelligence* 120 (2000), pp. 199–233. DOI: 10.1016/S0004-3702(00)00029-1. 25, 132, 160, 166, 173

[9] F. G. Cozman. 'Graphical models for imprecise probabilities'. *International Journal of Approximate Reasoning* 39.2-3 (2005), pp. 167–184. DOI: 10.1016/j.ijar.2004.10.003. 25

[10] F. G. Cozman & P. Walley. 'Graphoid properties of epistemic irrelevance and independence'. *Annals of Mathematics and Artificial Intelligence* 45 (1-2 2005), pp. 173–195. 14, 21, 189

[11] C. P. de Campos & F. G. Cozman. 'The inferential complexity of Bayesian and credal networks'. In: *Proceedings of the International Joint Conference on Artificial Intelligence*. Edinburgh, 2005, pp. 1313–1318. 190

[12] G. de Cooman & F. Hermans. 'Imprecise probability trees: bridging two theories of imprecise probability'. *Artificial Intelligence* 172 (2008), pp. 1400–1427. DOI: 10.1016/j.artint.2008.03.001. 13, 20, 21, 82, 114, 160, 187

[13] G. de Cooman, F. Hermans, A. Antonucci & M. Zaffalon. 'Epistemic irrelevance in credal nets: The case of imprecise Markov trees'. *International Journal of Approximate Reasoning* 51.9, SI (2010), pp. 1029–1052. DOI: 10.1016/j.ijar.2010.08.011. 13, 21, 161, 170

[14] G. de Cooman, F. Hermans & E. Quaeghebeur. 'Imprecise Markov chains and their limit behavior'. *Probability in the Engineering and Informational Sciences* 23.4 (2009), pp. 597–635. DOI: 10.1017/S0269964809990039. 14, 22, 135, 145, 146, 161

[15] G. de Cooman & E. Miranda. 'Symmetry of models versus models of symmetry'. In: *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.* Ed. by W. L. Harper & G. R. Wheeler. King's College Publications, 2007, pp. 67–149. 75, 76

[16] G. de Cooman & E. Miranda. 'Forward irrelevance'. *Journal of Statistical Planning and Inference* 139.2 (2009), pp. 256–276. DOI: 10.1016/j.jspi.2008.01.012. 112

[17] G. de Cooman, E. Miranda & M. Zaffalon. 'Independent natural ex- 42, 43, 49, 68, 72, 166

tension'. In: *Computational Intelligence for Knowledge-Based Systems. Lecture Notes in Artificial Intelligence 6178: proceedings of IPMU 2010*. Ed. by E. Hüllermeier, R. Kruse & F. Hoffman. Springer, 2010, pp. 737–746.

161 [18] G. de Cooman, E. Miranda & M. Zaffalon. 'Independent natural extension for sets of desirable gambles'. In: *ISIPTA'11: Proceedings of the Seventh International Symposium on Imprecise Probability: Theory and Applications*. Ed. by F. Coolen, G. de Cooman, T. Fetz & M. Oberguggenberger. Innsbruck: Studio Verlag, 2011, pp. 1–1.

161, 166–168 [19] G. de Cooman, E. Miranda & M. Zaffalon. 'Independent natural extension'. *Artificial Intelligence* 175.12-13 (2011), pp. 1911–1950. DOI: `10.1016/j.artint.2011.06.001`.

53, 76 [20] G. de Cooman & E. Quaeghebeur. 'Exchangeability and sets of desirable gambles'. *International Journal of Approximate Reasoning* 53.3 (2012), pp. 363 –395. DOI: `10.1016/j.ijar.2010.12.002`.

82 [21] G. de Cooman & M. C. M. Troffaes. *General Theory of Coherent Lower Previsions (To appear)*. Wiley Series in Probability and Statistics. New York: John Wiley & Sons, 2012.

24, 174 [22] G. de Cooman & M. Zaffalon. 'Updating beliefs with incomplete observations'. *Artificial Intelligence* 159.1-2 (2004), pp. 75–125. DOI: `10.1016/j.artint.2004.05.006`.

28, 202 [23] B. de Finetti. *Teoria delle Probabilità*. Turin: Einaudi, 1970.

11, 19, 24, 82, 127, 184, 186 [24] B. de Finetti. *Theory of Probability: A Critical Introductory Treatment*. English translation of [23], two volumes. Chichester: John Wiley & Sons, 1974–1975.

142 [25] E. V. Denardo. 'Periods of connected networks and powers of nonnegative matrices'. *Mathematics of Operations Research* 2.1 (1977), pp. 20–4. URL: `http://www.jstor.org/stable/3689120`.

80 [26] A. W. F. Edwards. 'Pascal's Problem: The 'Gambler's Ruin''. *International Statistical Review / Revue Internationale de Statistique* 51.1 (1983). History of Probability by Expectation, pp. 73–79. URL: `http://www.jstor.org/stable/1402732`.

122 [27] D. Harmanec. 'Generalizing Markov decision processes to imprecise

probabilities'. *Journal of Statistical Planning and Inference* 105 (2002), pp. 199–213. DOI: 10.1016/S0378-3758(01)00210-5.

[28] D. J. Hartfiel. *Markov Set-Chains*. Lecture Notes in Mathematics 1695. Berlin: Springer, 1998. 14, 22, 25, 122, 143

[29] F. Hermans & G. de Cooman. 'Characterisation of ergodic upper transition operators'. *International Journal of Approximate Reasoning* 53.4 (2012), pp. 573 –583. DOI: 10.1016/j.ijar.2011.12.008. 14, 22

[30] T. Herron, T. Seidenfeld & L. Wasserman. 'Divisive conditioning: further results on dilation'. *Philosophy of Science* 64 (1997), pp. 411–444. URL: http://www.jstor.org/stable/188318. 162, 183

[31] H. Itoh & K. Nakamura. 'Partially observable Markov decision processes with imprecise parameters'. *Artificial Intelligence* 171.8-9 (2007), pp. 453–490. DOI: 10.1016/j.artint.2007.03.004. 122

[32] G. F. J. & S. Rios. 'Quasi-Bayesian Behaviour: A more realistic approach to decision making?' *Trabajos de Estadística y de Investigación Operativa* 31.1 (1980), pp. 17–38. DOI: 10.1007/BF02888345. 28

[33] J. Jarvis & D. Shier. 'Graph-theoretic analysis of finite Markov chains'. In: *Applied Mathematical Modeling: A Multidisciplinary Approach*. Ed. by D. Shier & K. Wallenius. CRC Press, 1999, pp. 271–290. 192

[34] J. B. Kadane, M. J. Schervish & T. Seidenfeld. *Rethinking the foundations of statistics*. Cambridge studies in probability, induction, and decision theory. Cambridge University Press, 1999. 205

[35] J. G. Kemeny & J. L. Snell. *Finite Markov Chains*. Undergraduate Text in Mathematics. Reprint of the 1960 Edition. New York: Springer-Verlag, 1976. 125, 136, 137, 139, 145

[36] I. O. Kozine & L. V. Utkin. 'Interval-Valued Finite Markov Chains'. *Reliable Computing* 8.2 (2002), pp. 97–113. DOI: 10.1023/A:1014745904458. 25, 122

[37] B Lemmens & M Scheutzow. 'On the dynamics of sup-norm non-expansive maps'. *Ergodic Theory and Dynamical Systems* 25.Part 3 (2005), pp. 861–871. DOI: 10.1017/S0143385704000665. 123

[38] C. F. Manski. *Partial Identification of Probability Distributions*. New 24

York: Springer-Verlag, 2003.

161, 164 [39] E. Miranda. 'A survey of the theory of coherent lower previsions'. *International Journal of Approximate Reasoning* 48.2 (2008), pp. 628–658. DOI: 10.1016/j.ijar.2007.12.001.

161, 164 [40] E. Miranda. 'Updating coherent lower previsions on finite spaces'. *Fuzzy Sets and Systems* 160.9 (2009), pp. 1286–1307. DOI: 10.1016/j.fss.2008.10.005.

161, 169, 186 [41] E. Miranda & G. de Cooman. 'Marginal extension in the theory of coherent lower previsions'. *International Journal of Approximate Reasoning* 46.1 (2007), pp. 188–225. DOI: 10.1016/j.ijar.2006.12.009.

14, 21, 49, 189 [42] S. Moral. 'Epistemic irrelevance on sets of desirable gambles'. *Annals of Mathematics in Artificial Intelligence* 45.1-2 (2005), pp. 197–214. DOI: 10.1007/s10472-005-9011-0.

122 [43] A. Nilim & L. El Ghaoui. 'Robust control of Markov decision processes with uncertain transition matrices'. *Operations Research* 53 (2005), pp. 780–798. DOI: 10.1287/opre.1050.0216.

153 [44] R. Nussbaum. 'Hilberts projective metric and iterated nonlinear maps'. *Memoirs of the American Mathematical Society* 75.391 (1988), pp. 1–&.

123 [45] R. Nussbaum. 'Omega limit sets of nonexpansive maps: finiteness and cardinality estimates'. *Differential Integral Equations* 3.3 (1990), pp. 523–540.

123 [46] R. Nussbaum. 'Estimates of the periods of periodic points for nonexpansive operators'. *Israel Journal of Mathematics* 76 (3 1991), pp. 345–380. DOI: 10.1007/BF02773870.

14, 22, 173 [47] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, California: Morgan Kaufmann, 1988.

24, 82, 184 [48] F. P. Ramsey. 'Truth and probability (1926)'. In: *The Foundations of Mathematics and other Logical Essays*. Ed. by R. B. Braithwaite. London, 1931. Chap. VII, pp. 156–198.

122 [49] J. K. Satia & J. R. E. Lave. 'Markovian Decision Processes with Uncertain Transition Probabilities'. *Operations Research* 21.3 (1973), pp. 728–740. URL: http://www.jstor.org/stable/169381.

[50] T. Seidenfeld & L. Wasserman. 'Dilation for sets of probabilities'. *The Annals of Statistics* 21 (1993), pp. 1139–54. DOI: 10.1214/aos/1176349254. 162, 183

[51] T. Seidenfeld, M. J. Schervish & J. B. Kadane. 'A Representation of Partially Ordered Preferences'. *The Annals of Statistics* 23.6 (1995). Reprinted in [34], pp. 2168–2217. DOI: 10.1214/aos/1034713653. 28

[52] E. Seneta. *Non-negative Matrices and Markov Chains (Springer Series in Statistics)*. Springer, 2006. 143

[53] G. Shafer. *The Art of Causal Conjecture*. Cambridge, MA, USA: MIT Press, 1996. 13, 25, 81, 84–86, 88, 125, 186

[54] G. Shafer, P. R. Gillett & R. B. Scherl. 'A new understanding of subjective probability and its generalization to lower and upper prevision'. *International Journal of Approximate Reasoning* 33.1 (2003), pp. 1–49. DOI: 10.1016/S0888-613X(02)00134-2. 82

[55] G. Shafer & V. Vovk. *Probability and Finance: It's Only a Game!* New York: Wiley, 2001. 13, 21, 81, 82, 114

[56] R Sine. 'A Nonlinear Perron-Frobenius Theorem'. *Proceedings of the American Mathematical Society* 109.2 (1990), pp. 331–336. DOI: 10.2307/2047992. URL: http://www.jstor.org/stable/2047992. 123

[57] R. Tarjan. 'Depth-first search and linear graph algorithms'. *SIAM Journal on Computing* 1.2 (1972), pp. 146–60. DOI: 10.1137/0201010. 193

[58] V. Vovk. *Hoeffding's inequality in game-theoretic probability. The Game-Theoretic Probability and Finance project, Answer to Frequently Asked Question #3, http://probabilityandfinance.com*. 2007. 115

[59] D. Škulj. 'Finite discrete time Markov chains with interval probabilities'. In: *Soft Methods for Integrated Uncertainty Modelling*. Ed. by J. Lawry, E. Miranda, A. Bugarin, S. Li, M. A. Gil, P. Grzegorzewski & O. Hryniewicz. Berlin: Springer, 2006, pp. 299–306. 14, 22, 25, 122

[60] D. Škulj. 'Discrete time Markov chains with interval probabilities'. In: vol. 50. 8. 2009, pp. 1314 –1329. DOI: 10.1016/j.ijar.2009.06.007. 25, 122

[61] D. Škulj & R. Hable. 'Coefficients of Ergodicity for Imprecise Markov Chains'. In: *ISIPTA '09 – Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*. Ed. by 26, 153, 188

T. Augustin, F. P. Coolen, M. Serafin & M. C. Troffaes. SIPTA, 2009, pp. 377–386.

10, 12, 14, 18, 20, 21, 24, 28, 43, 45, 47, 53, 64, 68, 75, 76, 81, 108, 132, 161, 164, 169, 174, 184–186 [62] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Vol. 42. Monographs on Statistics and Applied Probability. London: Chapman & Hall, 1991.

.

190 [63] P. Walley. 'Inferences from multinomial data: Learning about a bag of marbles'. *Journal of the Royal Statistical Society, Series B* 58.1 (1996), pp. 3–34. URL: `http://www.jstor.org/stable/2346164`.

24, 184 [64] P. Walley. 'Measures of uncertainty in expert systems'. *Artificial Intelligence* 83.1 (1996), pp. 1–58. DOI: `10.1016/0004-3702(95)00009-7`.

122 [65] C. C. White & H. K. Eldeib. 'Markov decision-processes with imprecise transition-probabilities'. *Operations Research* 42 (1994), pp. 739–749. URL: `http://www.jstor.org/stable/171626`.

126, 127 [66] P. Whittle. *Probability via Expectation*. Fourth. New York: Springer, 2000.

12, 20, 28 [67] P. M. Williams. 'Indeterminate probabilities'. In: *Formal methods in the methodology of empirical sciences: Proceedings of the conference for formal methods in the methodology of empirical sciences*. Ed. by M. Przełęcki, K. Szaniawski & R. Wójcicki. Warsaw, 1974, pp. 229–246.

206 [68] P. M. Williams. *Notes on conditional previsions*. Tech. rep. Published as [69]. University of Sussex, 1975.

24, 82, 161, 184, 206 [69] P. M. Williams. 'Notes on Conditional Previsions'. *International Journal of Approximate Reasoning* 44 (2007). Original: [68], pp. 366–383. DOI: `10.1016/j.ijar.2006.07.019`.

174 [70] M. Zaffalon & E. Miranda. 'Conservative Inference Rule for Uncertain Reasoning under Incompleteness'. *Journal of Artificial Intelligence Research* 34 (2009), pp. 757–821.

Symbols

Number sets

$\mathbb{R}$ real numbers
$\mathbb{R}^*$ extended real numbers
$\mathbb{Q}$ rational numbers
$\mathbb{R}_{>0}$ positive real numbers
$\mathbb{R}_{<0}$ negative real numbers
$\mathbb{R}_{\neq 0}$ real numbers without zero
$\mathbb{R}_{\geq 0}$ nonneggative real numbers
$\mathbb{N}_{\geq 0}$ natural numbers (with zero)
$\mathbb{N}_{>0}$ natural numbers without zero

Mathematical symbols

$\Leftrightarrow$ is equivalent to
$\Rightarrow$ implies
$\wedge$ (logical) and
$\vee$ (logical) or
$\neg$ (logical) not
$\in$ is element of
$\notin$ is not an element of
$\subseteq$ is a subset of
$\subset$ is a strict subset of
$:=$ is by definition equal to
$=:$ which is the definition of
$>$ greater than
$<$ less than
$\geq$ greater or equal than
$\leq$ less or equal than
$\succ$ pointwise strictly greater
$\prec$ pointwise strictly smaller
A^c complement of the set A
$\mathcal{O}$ big O notation (complexity)
$|\cdot|$ absolute value or cardinality
$\lfloor \cdot \rfloor$ round to the largest integer smaller or equal than $\cdot$
$\lceil \cdot \rceil$ round to the smallest integer greater or equal to $\cdot$
$\|\cdot\|$ norm
$\|f\|_\infty = \sum_{x \in \mathcal{X}} |f(x)|$ supremum norm of f
$\|f\|_2 = \sum_{x \in \mathcal{X}} (f(x))^2$ Euclidean norm of f
$\|f\|_\text{v} = \sup f - \inf f$ variation

seminorm

$f|_A$ restriction of the function f to the space A, 27

int interior

cl closure

$\mathbb{1}$ identity matrix

$\omega_{\overline{\mathrm{T}}}(f)$ set of limit points of the trajectory $\overline{\mathrm{T}}^k f$, 123

Graph and tree symbols

$\mathcal{G}, \mathcal{H}$ any graph, 192

$\overline{\mathcal{G}}(\overline{\mathrm{T}})$ upper accessability graph used to check top class regularity,136

$\overline{\mathcal{G}}(\overline{\mathrm{T}})\big|_{\mathcal{R}}$ graph used to check top class regularity, 148

$\mathcal{G}^c$ Akian and Gaubert's critical graph, 154

$\mathcal{G}^f$ Akian and Gaubert's final graph, 154

$\overline{\mathcal{G}}(\overline{\mathrm{T}})$ upper accessibility graph used to check top class regularity, 148

$\sqsubseteq$ precedes, 83

$\sqsubset$ strictly precedes, 83

$\sqsupseteq$ follows, 83

$\sqsupset$ strictly follows, 83

$\parallel$ is not ordered with, 83

$\downarrow S$ nodes preceding the nodes in the set S, 83

$\uparrow S$ nodes preceding the nodes in the set S, 83

pa (s) parents of the node s, 162

mo (s) mother of the node s, 84

ch (s) children of the node s, 84

sib (s) siblings of the node s, 162

ancest (s) ancestors of the node s, 162

desc (s) descendants of the node s, 162

$\perp(s)$ nodes that are non-parents non-descendants of s, 165

depth(s) depth of s in a forward irrelevant process tree, 112

length$(\cdot)$ length of a path in a tree, 141

$\mathcal{C}_x$ set of simple cycles from the state x, 141

$\mathrm{var}_U F$ make a variable from the process F by evaluating in the cut U, 89

$\mathrm{Stop}_U F$ stop (keep constant) the process F in the cut U, 26

$\mathrm{Kill}_U F$ kill (make zero) the process F in the cut U, 26

$\square$ initial situation of a tree, 83

ω a generic terminal situation,

$\mathbb{S}$ set of all nodes/situations of a tree, 83

$\tilde{\mathbb{S}}$ trunk in a Markov tree, 178

$\mathbb{P}$ set of all paths, 85

$\mathbb{U}_s^V$ the set of all cuts following the situation s and preceding the cut V, 88

Ω set of terminal situations, 84

$E(a)$ (exact) event, 86

ω a generic terminal situation

$\mathbb{S}$ set of all nodes/situations of a tree, 83

$\tilde{\mathbb{S}}$ trunk in a Markov tree, 178

$\mathbb{P}$ set of all paths, 85

$\mathbb{U}_s^V$ the set of all cuts following the situation s and preceding the cut V, 88

Ω set of terminal situations, 84

$E(a)$ (exact) event corresponding to the situation a, 86

IP specific symbols

f, g generic name for a gamble,

h generic name for a (local) gamble,

I_A indicator function, 26

$\sqsubseteq$ is less resolved than, 38

$\sqsupseteq$ is more resolved than, 38

posi convex hull, 31

span finite linear span, 31

cccl cut-conglomerabiliy closure, 98

ext natural extension, 35

regext regular extension, 55

EI is epistemic irrelevant to, 69

$\otimes$ epistemic independent product, 70

$\boxtimes$ strong product, 166

$\mathrm{marg}_A(\cdot)$ marginalisation, 43

$\mathcal{A}|x$ condition on x, 47

$\mathcal{A}\rfloor x$ update on x, 47

$\mathrm{proj}_A f$ projection operator, 43

$\mathrm{proj}_A^T f$ lifted projection operator, 43

$\mathrm{cylext}_A f$ cylindrical extension, 45

$\mathrm{cylext}_A^T f$ lifted cylindrical extension, 45

$\mathrm{cylext}_A^{-1} f$ pre-image of lifted cylindrical extension, 45

$\mathcal{S}$ selection process, 92

$\sum_t \mathcal{S}$ summation of a selection process, 92

$\underline{M}$ submartingale, 114

$\overline{M}$ supermartingale, 114

ρ, ρ' coefficients of ergodicity, 153

Sets of gambles

$\mathcal{L}(\mathcal{X})$ Set of all gambles on $\mathcal{X}$ (linear space), 26

$\mathcal{N}$ Set of strictly desirable gambles, 56

$\mathcal{D}$ Set of desirable gambles, 30

$\mathcal{A}$ Set ofacceptable gambles, 29

$\mathcal{U}$ Set of undesirable gambles, 29

$\mathcal{I}$ Set of indifferent gambles, 30

$\mathcal{O}$ Set of unresolved gambles, 29

$\mathcal{R}$ Set of almost desirable gambles, 53

$\mathbb{A}$ Set of coherent sets of acceptable gambles, 34

$\mathbb{M}$ Set of all resolved sets of acceptable gambles, 39

$\mathcal{M}$ Set of resolved, acceptable gambles, 39

$\mathcal{P}$ Set of linear previsions (credal set), 132

$\mathcal{T}$ Set of linear transition operators, 132

$\mathrm{Sels}_t(\mathcal{A}_\cdot)$ Set of all t-selecions, 93

$\mathrm{SumSels}_t(\mathcal{A}_\cdot)$ Set of all summed t-selecions, 93

$\mathcal{A}_{E(\square)}$ Set of of acceptable gambles for event tree, 94

Previsions

p probability,

$\underline{p}$ lower probability,
$\overline{p}$ upper probability,
$\underline{q}$ local lower probability,
$\overline{q}$ local upper probability,
P linear prevision, 65
$\underline{\mathrm{P}}$ lower prevision, 60
$\overline{\mathrm{P}}$ upper prevision, 63
$\underline{\overline{\mathrm{P}}}$ lower-upper prevision
$\underline{\mathrm{N}}$ lower prevision under natural extension, 66
$\underline{\mathrm{R}}$ lower prevision under regular extension, 66
Q local linear prevision,
$\underline{\mathrm{Q}}$ local lower prevision,
$\overline{\mathrm{Q}}$ local upper prevision,
T linear transition operator, 126
$\underline{\mathrm{T}}$ lower transition operator, 130
$\overline{\mathrm{T}}$ upper transition operator, 129
$\mathbb{T}$ linear projection operator, 126
$\overline{\mathbb{T}}$ upper projection operator, 129
$\overline{P}_{xy}^{n}$ upper transition probability to go in n steps from x to y, 135