A link between Game-Theoretic Probability and Imprecise Probabilities

Gert de Cooman

Ghent University, SYSTeMS

gert.decooman@UGent.be
http://users.UGent.be/~gdcooma
gertekoo.wordpress.com

GTP2012
13 November 2012

My boon companions

FILIP HERMANS

ENRIQUE MIRANDA

JASPER DE BOCK

Imprecise probability models

Set of desirable gambles as a belief model

Two types of imprecise-probability models (Walley, 1991):

lower expectation: $\underline{P}(f(X))$ for all gambles $f\colon \mathscr{X} \to \mathbb{R}$

set of desirable gambles: $\mathscr{D} \subseteq \mathscr{G}(\mathscr{X})$ is a set of gambles that a subject strictly prefers to zero

Set of desirable gambles as a belief model

Two types of imprecise-probability models (Walley, 1991):

lower expectation: $\underline{P}(f(X))$ for all gambles $f\colon \mathscr{X} \to \mathbb{R}$

set of desirable gambles: $\mathscr{D} \subseteq \mathscr{G}(\mathscr{X})$ is a set of gambles that a subject strictly prefers to zero

Working with sets of desirable gambles $\mathscr{D}$:

- is simpler, more intuitive and more elegant
- is more general and expressive than (conditional) lower expectations and even full conditional measures
- gives a geometrical flavour to probabilistic inference
- shows that probabilistic inference is 'logical' inference
- avoids problems with conditioning on sets of probability zero

Coherence for a set of desirable gambles

A set of desirable gambles $\mathscr{D}$ is called coherent if:

D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]

Coherence for a set of desirable gambles

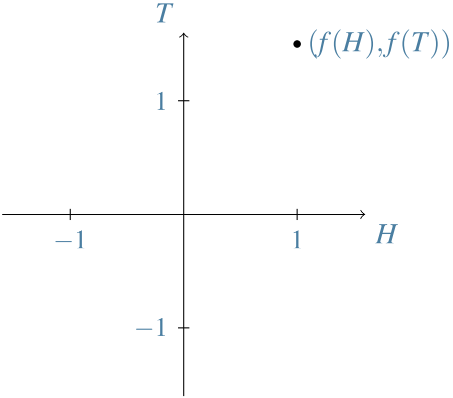
A set of desirable gambles $\mathscr{D}$ is called coherent if:

D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]



Coherence for a set of desirable gambles

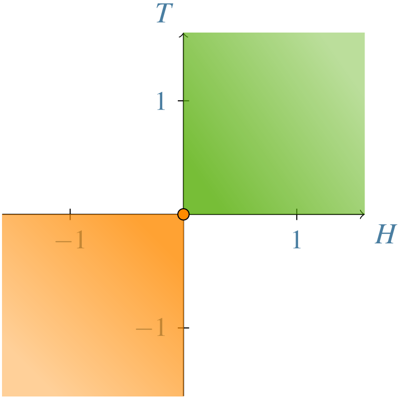
A set of desirable gambles $\mathscr{D}$ is called coherent if:

D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]



Coherence for a set of desirable gambles

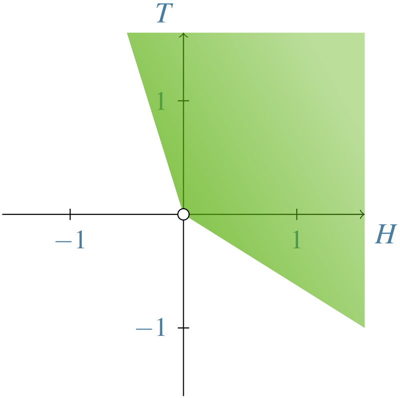
A set of desirable gambles $\mathscr{D}$ is called coherent if:

D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]



Coherence for a set of desirable gambles

A set of desirable gambles $\mathscr{D}$ is called coherent if:

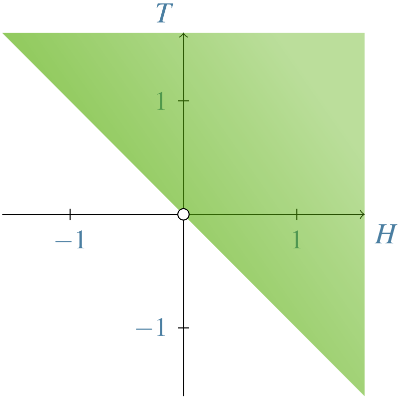
D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]

Precise models correspond to the special case that the convex cones $\mathscr{D}$ are actually halfspaces!



Coherence for a set of desirable gambles

A set of desirable gambles $\mathscr{D}$ is called coherent if:

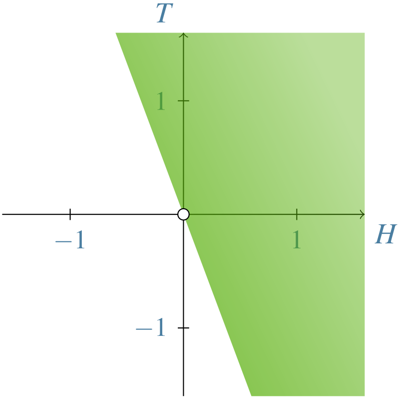
D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

D2. if $f > 0$ then $f \in \mathscr{D}$ [desiring partial gains]

D3. if $f, g \in \mathscr{D}$ then $f + g \in \mathscr{D}$ [addition]

D4. if $f \in \mathscr{D}$ then $\lambda f \in \mathscr{D}$ for all real $\lambda > 0$ [scaling]

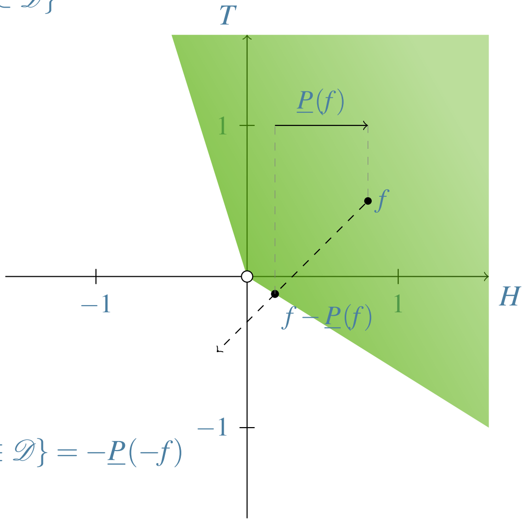
Precise models correspond to the special case that the convex cones $\mathscr{D}$ are actually halfspaces!



Connection with lower and upper expectations

lower expectation:

$\underline{P}(f) = \sup\{\alpha \in \mathbb{R} : f - \alpha \in \mathscr{D}\}$

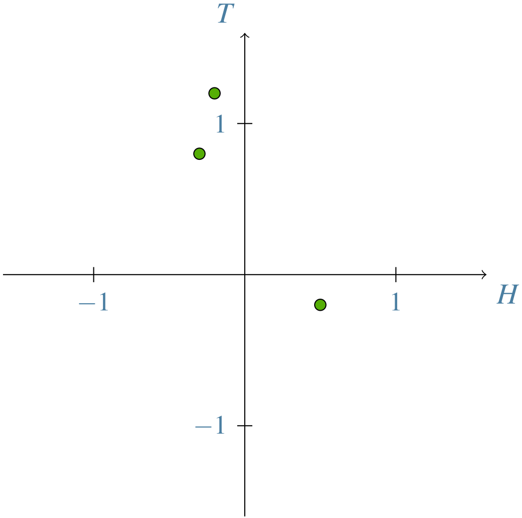


upper expectation:

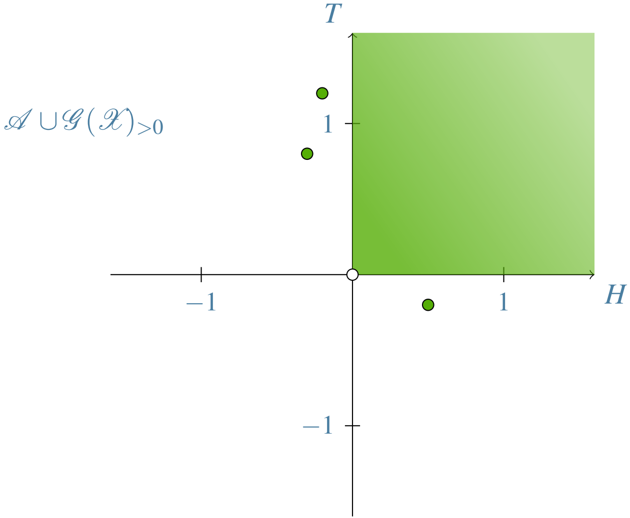
$\overline{P}(f) = \inf\{\alpha \in \mathbb{R} : \alpha - f \in \mathscr{D}\} = -\underline{P}(-f)$

Inference: natural extension

$\mathscr{A}$

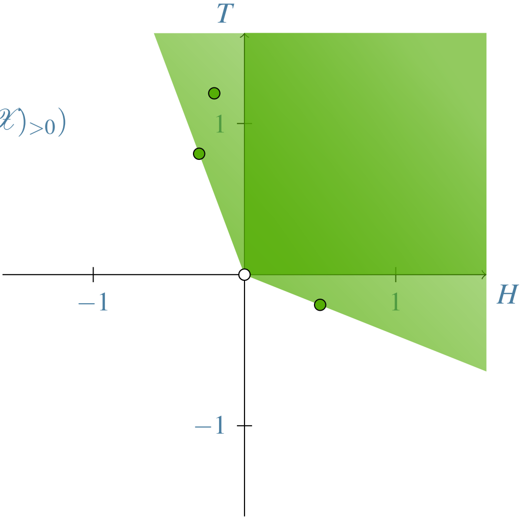


Inference: natural extension



Inference: natural extension

$$\mathscr{E}_{\mathscr{A}} := \operatorname{posi}(\mathscr{A} \cup \mathscr{G}(\mathscr{X})_{>0})$$



$$\operatorname{posi}(\mathscr{K}) := \left\{ \sum_{k=1}^{n} \lambda_k f_k : f_k \in \mathscr{K}, \lambda_k > 0, n > 0 \right\}$$

The conditioning rule

Now suppose that you learn that B occurs.

This leads to an updated set of desirable gambles:

$$f \in \mathscr{D}\rfloor B \Leftrightarrow I_B f \in \mathscr{D} \text{ or } f > 0$$

or equivalently, for gambles g on B:

$$g \in \mathscr{D}\rfloor B \Leftrightarrow I_B g \in \mathscr{D}.$$

The conditioning rule

Now suppose that you learn that B occurs.

This leads to an updated set of desirable gambles:

$$f \in \mathscr{D}|B \Leftrightarrow I_B f \in \mathscr{D} \text{ or } f > 0$$

or equivalently, for gambles g on B:

$$g \in \mathscr{D}\rfloor B \Leftrightarrow I_B g \in \mathscr{D}.$$

Just like in the unconditional case, we can use a coherent set of desirable gambles $\mathscr{D}$ to derive conditional lower and upper expectations:

$$\underline{P}(f|B) := \sup\{\alpha \in \mathbb{R} \colon f - \alpha \in \mathscr{D}|B\} = \sup\{\alpha \in \mathbb{R} \colon I_B(f - \alpha) \in \mathscr{D}\}$$
$$\underline{P}(g|B) := \sup\{\alpha \in \mathbb{R} \colon g - \alpha \in \mathscr{D}\rfloor B\} = \sup\{\alpha \in \mathbb{R} \colon I_B(g - \alpha) \in \mathscr{D}\}$$

All you know about probability theory . . .

All of propositional logic and probability theory can be inferred from:

- the coherence rules D1–D4
- the conditioning rule
- (and some extra continuity requirements)

for sets of desirable gambles.

1. Bayes's Rule and Theorem
2. laws of large numbers
3. other limit laws
4. . . .

But they provide a solid foundation for imprecise probabilities too!

Grafting IP-models on an event tree

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Artificial Intelligence 172 (2008) 1400–1427

Artificial Intelligence

www.elsevier.com/locate/artint

Imprecise probability trees: Bridging two theories of imprecise probability

Gert de Cooman *, Filip Hermans

Ghent University, SYSTeMS Research Group, Technologiepark – Zwijnaarde 914, 9052 Zwijnaarde, Belgium

Received 30 March 2007; received in revised form 12 February 2008; accepted 3 March 2008

Available online 18 March 2008

Abstract

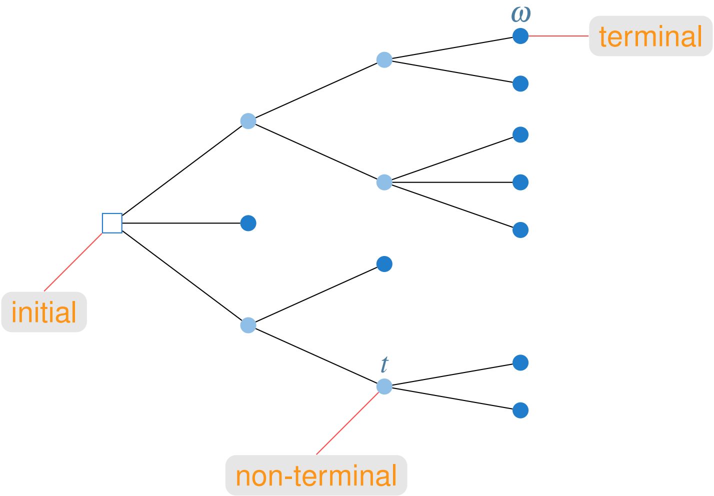
We give an overview of two approaches to probability theory where lower and upper probabilities, rather than probabilities, are used: Walley's behavioural theory of imprecise probabilities, and Shafer and Vovk's game-theoretic account of probability. We show that the two theories are more closely related than would be suspected at first sight, and we establish a correspondence between them that (i) has an interesting interpretation, and (ii) allows us to freely import results from one theory into the other. Our approach leads to an account of probability trees and random processes in the framework of Walley's theory. We indicate how our results can be used to reduce the computational complexity of dealing with imprecision in probability trees, and we prove an interesting and quite general version of the weak law of large numbers.

Keywords: Game-theoretic probability; Imprecise probabilities; Coherence; Conglomerability; Event tree; Probability tree; Imprecise probability tree; Lower prevision; Immediate prediction; Prequential Principle; Law of large numbers; Hoeffding's inequality; Markov chain; Random process

```
@ARTICLE{cooman2008,
   author = {de Cooman, Gert and Hermans, Filip},
   title = {Imprecise probability trees: Bridging two theories of imprecise probability},
   journal = {Artificial Intelligence},
   year = {2008},
   volume = {172},
   pages = {1400–1427},
   number = {11},
   doi = {10.1016/j.artint.2008.03.001}
}
```

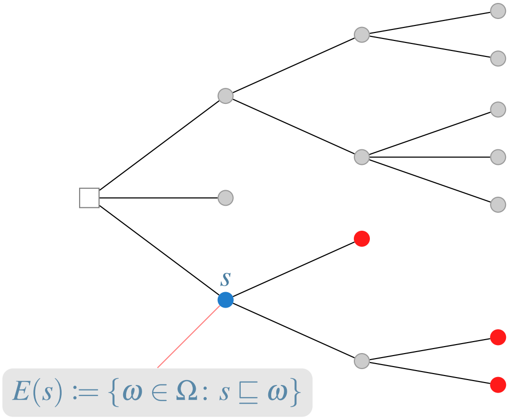
An event tree and its situations

Situations are nodes in the event tree, and the sample space Ω is the set of all terminal situations:

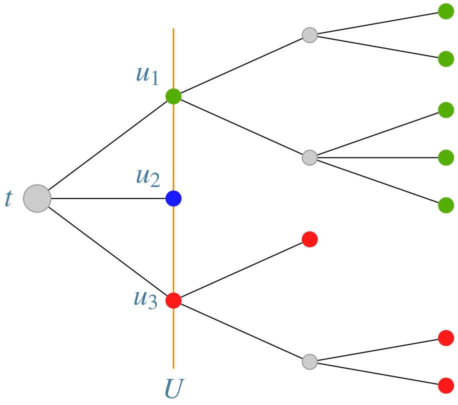


Events

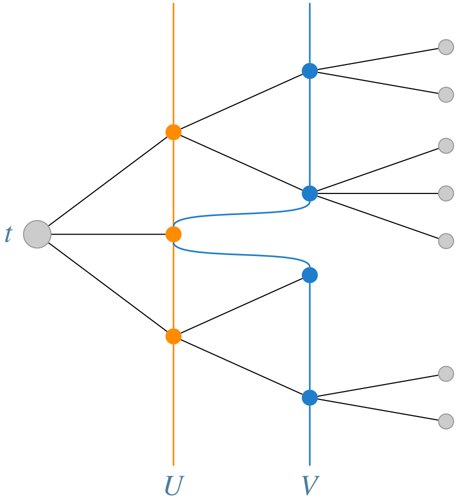
An event A is a subset of the sample space Ω:



Cuts of a situation



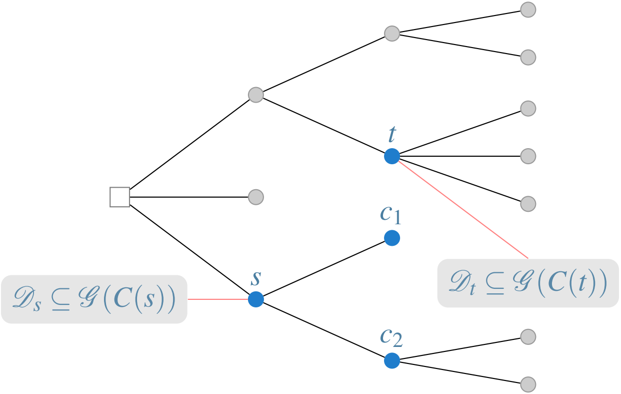
Cuts of a situation



U precedes V: $U \sqsubseteq V$

Immediate prediction models

In each non-terminal situation s, Forecaster has a belief model $\mathscr{D}_s$, satisfying D1–D4.



$C(s) = \{c_1, c_2\}$ is the set of children of s.

Imprecise probability trees with bounded horizon

From a local to a global model

We first assume that the event tree has bounded depth.

How to combine the local pieces of information into a coherent global model:

Forecaster accepts which gambles f on the entire sample space Ω?

For each non-terminal situation s and each $h_s \in \mathscr{D}_s$, Forecaster accepts the gamble $\hat{h}_s$ on Ω, where

$$\hat{h}_s(\omega) := \begin{cases} 0 & \omega \notin E(s) \\ h_s(c_\omega) & s \sqsubset c_\omega \sqsubseteq \omega. \end{cases}$$

$\hat{h}_s$ represents the gamble on Ω that is called off unless Reality ends up in situation s, and then depends only on Reality's move c immediately after s, and gives the same value $h_s(c)$ to all paths ω that go through c.

Natural extension

So Forecaster accepts all gambles in the set:

$$\mathscr{D} := \{\hat{h}_s : h_s \in \mathscr{D}_s \text{ and } s \text{ non-terminal}\}.$$

Find the natural extension $\mathscr{E}_{\mathscr{D}}$ of $\mathscr{D}$:

the smallest subset of $\mathscr{G}(\Omega)$ that includes $\mathscr{D}$, is coherent—satisfies D1-D4—and satisfies cut conglomerability.

A set of desirable gambles $\mathscr{D}$ on Ω is cut-conglomerable if for all cuts U of $\square$:

D5. if $(\forall u \in U)(I_{E(u)} f \in \mathscr{D} \cup \{0\})$ then $f \in \mathscr{D} \cup \{0\}$.

Desirable selections and gamble processes

A desirable t-selection $\mathscr{S}$ is a process, defined on all non-terminal situations s that follow t, and such that

$$\mathscr{S}(s) \in \mathscr{D}_s \cup \{0\} \text{ for all non-terminal } s \sqsupseteq t$$
$$\mathscr{S}(s) \neq 0 \text{ for some non-terminal } s \sqsupseteq t$$

It selects, in advance, a desirable-or-zero gamble $\mathscr{S}(s)$ from the available desirable gambles in each non-terminal $s \sqsupseteq t$.

With a desirable t-selection $\mathscr{S}$, we can associate a real-valued t-gamble process $\mathscr{I}^{\mathscr{S}}$, which is a t-process such that:

$$\mathscr{I}^{\mathscr{S}}(c) := \mathscr{I}^{\mathscr{S}}(s) + \mathscr{S}(s)(c), \text{ for all } s \sqsupseteq t \text{ and all } c \in C(s)$$

and

$$\mathscr{I}^{\mathscr{S}}(t) = 0.$$

Marginal Extension Theorem

Theorem (Marginal Extension Theorem)

There is a smallest set of gambles that satisfies D1–D4 and D5 and that includes $\mathscr{D}$. This natural extension of $\mathscr{D}$ is given by

$$\mathscr{E}_{\mathscr{D}} := \left\{ g \in \mathscr{G}(\Omega) \colon g \geq \mathscr{I}_{\Omega}^{\mathscr{S}} \text{ for some desirable } \square\text{-selection } \mathscr{S} \right\}.$$

Moreover, for any non-terminal situation t and any t-gamble g, $I_{E(t)} g \in \mathscr{E}_{\mathscr{D}}$ iff there is some desirable t-selection $\mathscr{S}_t$ such that $g \geq \mathscr{I}_{E(t)}^{\mathscr{S}_t}$.

Use the coherent set of desirable gambles $\mathscr{E}_{\mathscr{D}}$ to define predictive lower expectations $\underline{P}(\cdot|t) := \underline{P}(\cdot|E(t))$ conditional on an event $E(t)$:

For any t-gamble f on $E(t)$ and for any non-terminal situation t,

$$\begin{aligned}\underline{P}(f|t) &:= \sup \left\{ \alpha \in \mathbb{R} \colon I_{E(t)}(f - \alpha) \in \mathscr{E}_{\mathscr{D}} \right\} \\ &= \sup \left\{ \alpha \in \mathbb{R} \colon f - \alpha \geq \mathscr{I}_{E(t)}^{\mathscr{S}} \text{ for some desirable } t\text{-selection } \mathscr{S} \right\}.\end{aligned}$$

Law of Iterated Expectations

For a cut U of t, define the t-gamble $\underline{P}(f|U)$ by

$$\underline{P}(f|U)(\omega) := \underline{P}(f|u) \text{ for the unique } u \in U \text{ such that } u \sqsubseteq \omega.$$

Theorem (Law of Iterated Expectations)

Consider any two (bounded) cuts U and V of a situation t such that U precedes V. Then for all t-gambles f on $E(t)$,

1. $\underline{P}(f|t) = \underline{P}(\underline{P}(f|U)|t)$;
2. $\underline{P}(f|U) = \underline{P}(\underline{P}(f|V)|U)$.

Applications: stochastic processes

Probability in the Engineering and Informational Sciences, **23**, 2009, 597–635. Printed in the U.S.A.
doi:10.1017/S0269964809990039

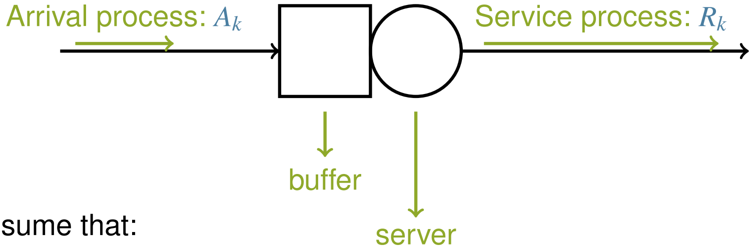
IMPRECISE MARKOV CHAINS AND THEIR LIMIT BEHAVIOR

GERT DE COOMAN, FILIP HERMANS, AND ERIK QUAEGHEBEUR
SYSTeMS Research Group
Ghent University
Technologiepark–Zwijnaarde 914, 9052 Zwijnaarde, Belgium
E-mail: {gert.decooman, filip.hermans, erik.quaeghebeur}@ugent.be

When the initial and transition probabilities of a finite Markov chain in discrete time are not well known, we should perform a sensitivity analysis. This can be done by considering as basic uncertainty models the so-called *credal sets* that these probabilities are known or believed to belong to and by allowing the probabilities to vary over such sets. This leads to the definition of an *imprecise Markov chain*. We show that the time evolution of such a system can be studied very efficiently using so-called *lower* and *upper expectations*, which are equivalent mathematical representations of credal sets. We also study how the inferred credal set about the state at time n evolves as $n \to \infty$: under quite unrestrictive conditions, it converges to a uniquely invariant credal set, regardless of the credal set given for the initial state. This leads to a non-trivial generalization of the classical Perron–Frobenius theorem to imprecise Markov chains.

```
@ARTICLE{cooman2009,
  author = {{d}e Cooman, Gert and Hermans, Filip and Quaegehebeur, Erik},
  title = {Imprecise {M}arkov chains and their limit behaviour},
  journal = {Probability in the Engineering and Informational Sciences},
  year = 2009,
  volume = 23,
  pages = {597--635},
  doi = {10.1017/S0269964809990039}
}
```

Queueing system

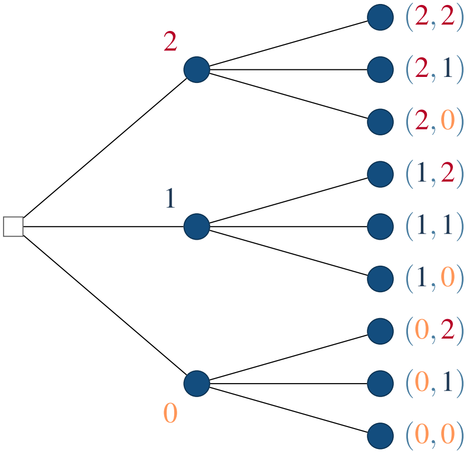


We assume that:

- There is only one queue
- There is only one server
- The capacity of the queueing system is 2
- There is maximally one arrival in one time step
- There is maximally one item serviced in one time step
- The service decision happens before the arrival event

Unrolling the event tree

The state of the system X_k at time k is an element of $\{0, 1, 2\}$ where 0 corresponds to $X_k = 0$, 1 to $X_k = 1$ and 2 corresponds to $X_k = 2$.



The relation between X_k, A_k and R_k

The number of objects in the system X_{k+1} (= in the buffer + being serviced) at time $k+1$, is determined by:

- X_k: The number of objects in the system at time k,
- A_k: The number of objects that arrive at time k,
- R_k: The number of objects that are serviced at time k,

$$X_{k+1} = X_k + A_k - R_k.$$

Only a limited number of combinations of A_k, R_k, X_k and X_{k+1} are allowed:

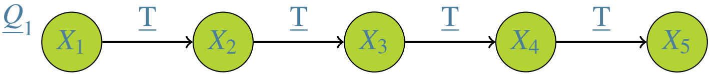
X_{k+1}	X_k	A_k	R_k
0	0	0	0
	~~0~~	~~0~~	~~1~~
1	0	1	0
	~~0~~	~~1~~	~~1~~

X_{k+1}	X_k	A_k	R_k
1	1	0	0
1	1	0	1
2	1	1	0
1	1	1	1

X_{k+1}	X_k	A_k	R_k
2	2	0	0
1	2	0	1
	~~2~~	~~1~~	~~0~~
2	2	1	1

Imprecise (stationary) Markov chain

- We assume that A_k and R_k do not depend on $A_{1:k-1}, R_{1:k-1}$.
- Consequently, X_k is independent of $X_{1:k-1}$ which is the Markov condition and $\{X_k\}_{k\in\mathbb{N}}$ is a discrete time, imprecise Markov chain.
- We assume that the belief model for (A_k, R_k) does not depend on the time index k: the resulting Markov chain is stationary.

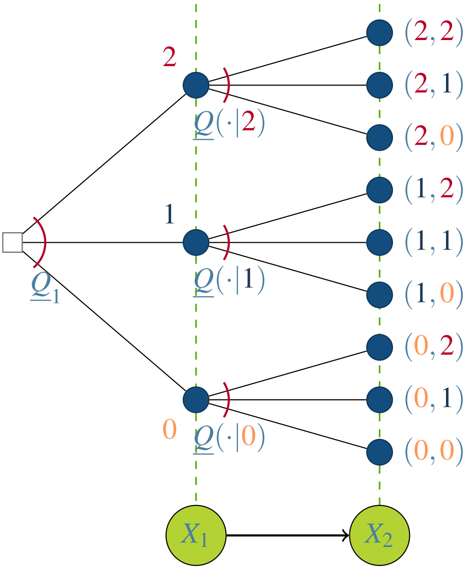


An imprecise stationary Markov chain is defined by

- its state space $\mathscr{X}$,
- the prior belief model $\underline{Q}_1$ on $\mathscr{G}(\mathscr{X})$,
- the upper transition operator $\underline{\mathrm{T}}\colon \mathscr{G}(\mathscr{X}) \to \mathscr{G}(\mathscr{X})$

$$\underline{\mathrm{T}}f(x) := \underline{Q}(f|x) \text{ for all states } x \in \mathscr{X}.$$

Markov chains are a special type of event tree



- In each situation, there is the choice of the same possibilities $\mathscr{X} = \{1, 0, 2\}$. This is what we call the state space.
- The belief model depends only on the last state. This is the Markov condition

Law of iterated expectations for Markov chains

The advantage of interpreting the queueing system as an imprecise Markov chain is that any expectation (assuming epistemic irrelevance in the Markov condition) can be calculated recursively.

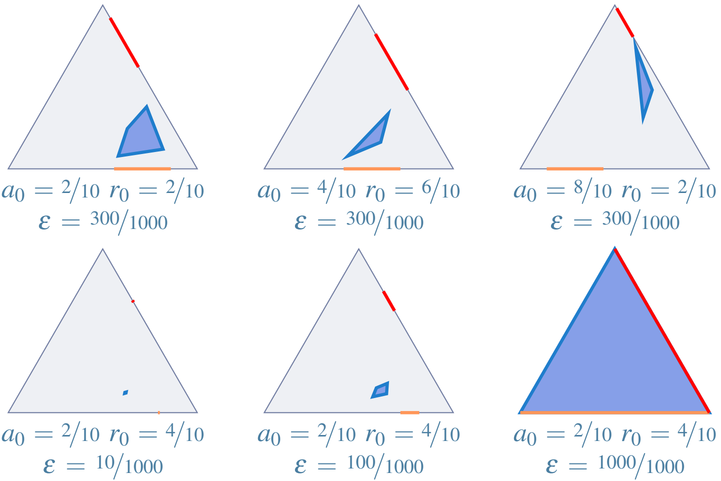
Theorem

For any real-valued map h on $\mathscr{X}_n$, and for any $1 \leq \ell < n$ and all x_ℓ in $\mathscr{X}_\ell$:

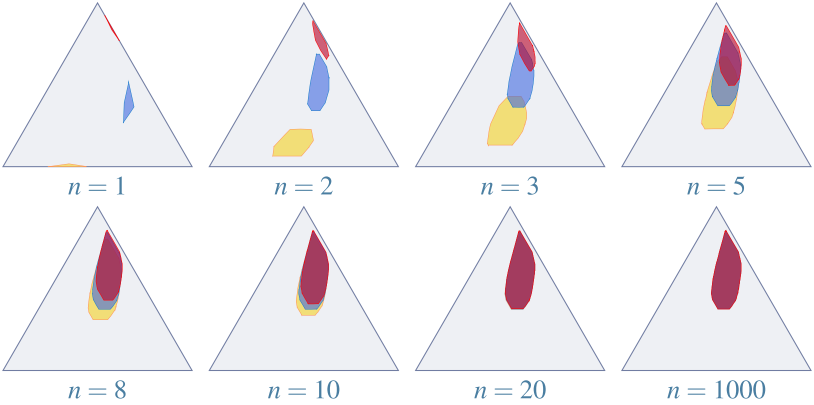
$$\underline{P}_{n|\ell}(h|x_\ell) = \underline{\mathrm{T}}^{n-\ell} h(x_\ell),$$

$$\underline{P}_n(h) = \underline{Q}_1(\underline{\mathrm{T}}^{n-1} h).$$

The transition operator on the simplex

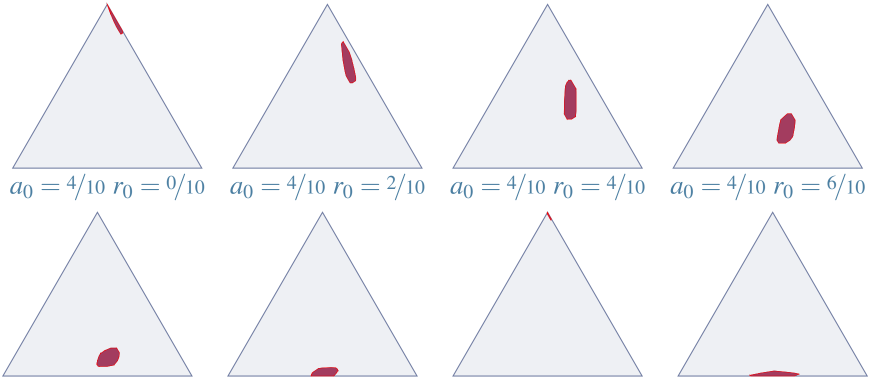


Time evolution and ergodicity



$a_0 = 7/10$ and $r_0 = 4/10$ and $\varepsilon = 2/10$.

Influence of the precise models

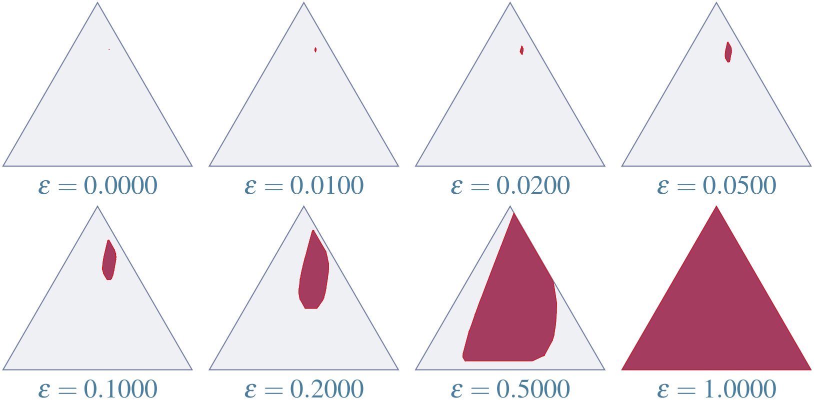


$a_0 = {}^4/_{10}\ r_0 = {}^0/_{10}$ $a_0 = {}^4/_{10}\ r_0 = {}^2/_{10}$ $a_0 = {}^4/_{10}\ r_0 = {}^4/_{10}$ $a_0 = {}^4/_{10}\ r_0 = {}^6/_{10}$

$a_0 = {}^4/_{10}\ r_0 = {}^8/_{10}$ $a_0 = {}^4/_{10}\ r_0 = {}^{10}/_{10}$ $a_0 = {}^8/_{10}\ r_0 = {}^0/_{10}$ $a_0 = {}^2/_{10}\ r_0 = {}^{10}/_{10}$

$n = 5000$ and $\varepsilon = {}^1/_{10}$.

Influence of imprecision



$n = 5000$ and $a_0 = {}^{7}/_{10}$ and $r_0 = {}^{4}/_{10}$.

The Perron-Frobenius theorem

Theorem (Perron–Frobenius Theorem)

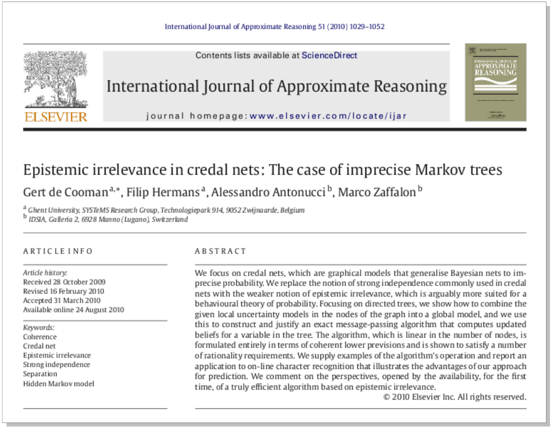
Consider a stationary imprecise Markov chain with finite state set $\mathscr{X}$ *that is ergodic. Then for every initial upper expectation* $\underline{P}_1$*, the upper expectation* $\underline{P}_n = \underline{P}_1 \circ \underline{\mathrm{T}}^{n-1}$ *for the state at time* n *converges point-wise to the same upper expectation* $\underline{P}_\infty$*:*

$$\lim_{n\to\infty} \underline{P}_n(h) = \lim_{n\to\infty} \underline{P}_1(\underline{\mathrm{T}}^{n-1}h) =: \underline{P}_\infty(h) \text{ for all } h \text{ in } \mathscr{G}(\mathscr{X}).$$

Moreover, the limit upper expectation $\underline{P}_\infty$ *is the only* $\underline{\mathrm{T}}$*-invariant upper expectation on* $\mathscr{G}(\mathscr{X})$*.*

Applications: credal trees



International Journal of Approximate Reasoning 51 (2010) 1029–1052

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Epistemic irrelevance in credal nets: The case of imprecise Markov trees

Gert de Cooman [a,*], Filip Hermans [a], Alessandro Antonucci [b], Marco Zaffalon [b]

[a] Ghent University, SYSTeMS Research Group, Technologiepark 914, 9052 Zwijnaarde, Belgium
[b] IDSIA, Galleria 2, 6928 Manno (Lugano), Switzerland

ARTICLE INFO

Article history:
Received 28 October 2009
Revised 16 February 2010
Accepted 31 March 2010
Available online 24 August 2010

Keywords:
Coherence
Credal net
Epistemic irrelevance
Strong independence
Separation
Hidden Markov model

ABSTRACT

We focus on credal nets, which are graphical models that generalise Bayesian nets to imprecise probability. We replace the notion of strong independence commonly used in credal nets with the weaker notion of epistemic irrelevance, which is arguably more suited for a behavioural theory of probability. Focusing on directed 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. The algorithm, which is linear in the number of nodes, is formulated entirely in terms of coherent lower previsions and is shown to satisfy a number of rationality requirements. We supply examples of the algorithm's operation and report an application to on-line character recognition that illustrates the advantages of our approach for prediction. We comment on the perspectives, opened by the availability, for the first time, of a truly efficient algorithm based on epistemic irrelevance.


```
@ARTICLE{cooman2010,
   author = {{d}e Cooman, Gert and Hermans, Filip and Antonucci, Alessandro and Zaffalon, Marco},
   title = {Epistemic irrelevance in credal nets: the case of imprecise {M}arkov trees},
   journal = {International Journal of Approximate Reasoning},
   year = 2010,
   volume = 51,
   pages = {1029--1052},
   doi = {10.1016/j.ijar.2010.08.011}
}
```

Credal trees: local uncertainty models

Local uncertainty model associated with each node t

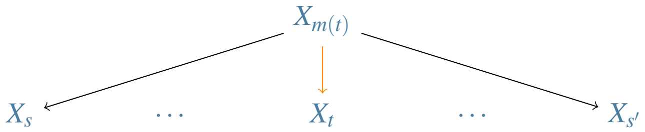
For each possible value $x_{m(t)} \in \mathscr{X}_{m(t)}$ of the mother variable $X_{m(t)}$, we have a conditional lower expectation

$$\underline{Q}_t(\cdot|x_{m(t)}) \colon \mathscr{G}(\mathscr{X}_t) \to \mathbb{R}$$

where

$$\underline{Q}_t(f|x_{m(t)}) = \text{lower expectation of } f(X_t)\text{, given that } X_{m(t)} = x_{m(t)}.$$

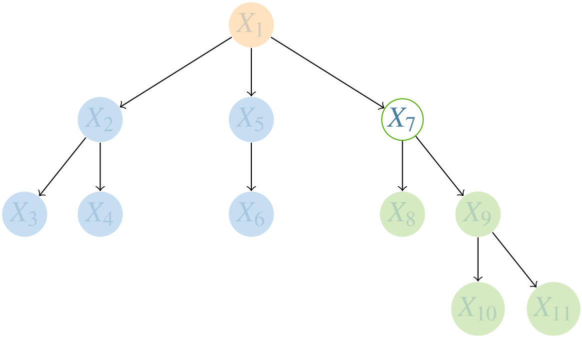
The local model $\underline{Q}_t(\cdot|X_{m(t)})$ is a conditional lower expectation operator.



Interpretation of the graphical structure

The graphical structure is interpreted as follows:

Conditional on the mother variable, the non-parent non-descendants of each node variable are *epistemically irrelevant* to it and its descendants.



MePICTIr for updating a credal tree

For a credal tree we can find the joint model from the local models recursively, from leaves to root.

Exact message passing algorithm

- credal tree treated as an expert system
- linear complexity in the number of nodes

Python code

- written by Filip Hermans
- testing and connection with strong independence in cooperation with Marco Zaffalon and Alessandro Antonucci

Current (toy) applications in HMMs

character recognition, air traffic trajectory tracking and identification, earthquake rate prediction

7th International Symposium on Imprecise Probability: Theories and Applications, Innsbruck, Austria, 2011

State sequence prediction in imprecise hidden Markov models

Jasper De Bock and Gert de Cooman
SYSTeMS, Ghent University, Belgium
{jasper.debock,gert.decooman}@UGent.be

Abstract

We present an efficient exact algorithm for estimating state sequences from outputs (or observations) in imprecise hidden Markov models (iHMM), where both the uncertainty linking one state to the next, and that linking a state to its output, are represented using coherent lower previsions. The notion of independence we associate with the credal network representing the iHMM is that of epistemic irrelevance. We consider as best estimates for state sequences the (Walley–Sen) maximal sequences for the posterior joint state model (conditioned on the observed output sequence), associated with a gain function that is the indicator of the state sequence. This corresponds to (and generalises) finding the state sequence with the highest posterior probability in HMMs with precise transition and output probabilities (pHMMs). We argue that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and essentially linear in the number of maximal state sequences. For binary iHMMs, we investigate experimentally how the number of maximal state sequences depends on the model parameters.

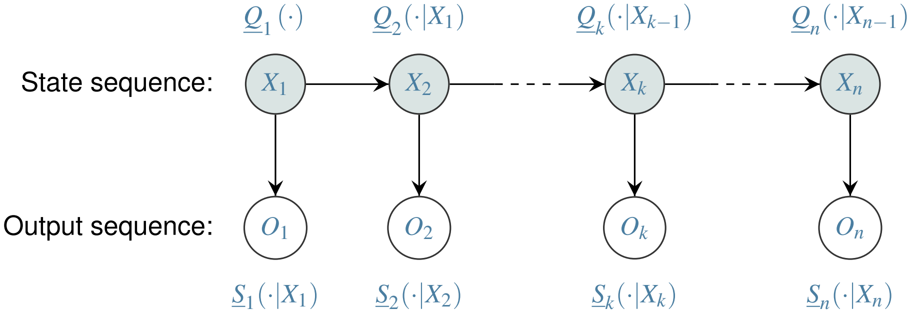
Keywords. Imprecise hidden Markov model, optimal state sequence, maximality, coherent lower prevision, credal network, epistemic irrelevance.

is a serious limitation, there are, nevertheless quite a number of models and applications that involve a tree structure. Amongst these, hidden Markov models (HMMs) are definitely the simplest, and perhaps also the most popular ones. But this brings us to the second limitation: MePiCTIr only allows updating of beliefs about a *single* node. Whereas one of the most important applications for, say, HMMs, involves finding the *sequence* of (hidden) states with the highest posterior probability after observing a sequence of outputs [11]. For HMMs with precise local transition and emission probabilities, there are quite efficient dynamic programming algorithms, such as Viterbi's [11, 13], for performing this task. For imprecise-probabilistic local models, such as coherent lower previsions, we know of no algorithm in the literature for which the computational complexity comes even close to that of Viterbi's.

In this paper, we take the first steps towards remedying this situation. We describe imprecise hidden Markov models as special cases of credal trees (a special case of credal networks) under epistemic irrelevance in Section 2. We show in particular how we can use the ideas underlying the MePiCTIr algorithm (independent natural extension and marginal extension) to construct a most conservative joint model from imprecise local transition and emission models, and derive a number of interesting and useful formulas from that construction. In Section 3 we explain how a sequence

```
@INPROCEEDINGS{cooman2011,
  author = {De Bock, Jasper and {d}e Cooman, Gert},
  title = {State sequence prediction in imprecise hidden Markov models},
  booktitle = {ISIPTA '11 -- Proceedings of the 7th International Symposium on Imprecise Probability: Theories and Applications},
  year = 2011,
  editor = {Coolen, Frank P. A. and {d}e Cooman, Gert and Fetz, Thomas and Oberguggenberger, Michael},
  address = {Innsbruck, Austria},
  publisher = {SIPTA}
}
```

A HMM is a special credal tree



Maximal state sequences

Classically (Viterbi):

Find the state sequence $\hat{x}_{1:n}$ that maximises the posterior probability $p(x_{1:n}|o_{1:n})$ corresponding to a given observation sequence $o_{1:n}$.

Maximality (under robust ordering):

Define a partial order $>$ on state sequences:

$$\hat{x}_{1:n} > x_{1:n} \text{ iff } p(\hat{x}_{1:n}|o_{1:n}) > p(x_{1:n}|o_{1:n}) \text{ for all compatible } p(\cdot|o_{1:n})$$

Find the state sequences $\hat{x}_{1:n}$ that are maximal: undominated by any other state sequence.

ESTIHMM for finding all maximal state sequences

Exact backward-forward algorithm

- developed by Jasper De Bock
- finds all maximal state sequences that correspond to a given observation sequence
- quadratic complexity in the number of nodes [linear]
- cubic complexity in the number of states [quadratic]
- linear complexity in the number of maximal sequences. [linear]

Python code

- written by Jasper De Bock

Current (toy) applications in HMMs

character recognition, finding gene islands

Imprecise probability trees with unbounded horizon

What we would like to get to

We now allow the discrete tree to have unbounded depth.

Define for any t-process $\mathscr{F}$ the t-gamble $\limsup \mathscr{F}$ as:

$$\limsup \mathscr{F}(\omega) := \limsup_{n\to+\infty} \mathscr{F}(\omega_n) \text{ for all } \omega \in E(t),$$

where ω_n denotes the (finite or denumerably infinite) sequence of situations in the path ω.

We would like to go from

$$\underline{P}(f|t) = \sup\left\{\alpha : f - \alpha \geq \mathscr{I}_{E(t)}^{\mathscr{S}} \text{ for some desirable } t\text{-selection } \mathscr{S}\right\}$$

to

$$\underline{P}(f|t) = \sup\left\{\alpha : f - \alpha \geq \limsup \mathscr{I}^{\mathscr{S}} \text{ for some desirable } t\text{-selection } \mathscr{S}\right\}.$$

This is the counterpart of the Shafer–Vovk–Ville formula.

Additional axioms

This seems impossible with only D1–D4 (coherence) and D5 (cut conglomerability).

So we add two axioms to coherence: bounded cut conglomerability

D5. For all bounded cuts U of $\Box$:

$$(\forall u \in U)(I_{E(u)} f \in \mathscr{D} \cup \{0\}) \Rightarrow f \in \mathscr{D} \cup \{0\}.$$

and bounded cut continuity

D6. For any real process $\mathscr{F}$ such that $\limsup_{U \text{ bounded}} \mathscr{F}_U \in \mathscr{G}(\Omega)$, and such that $\mathscr{F}_V - \mathscr{F}_U \in \mathscr{D} \cup \{0\}$ for all bounded cuts $U \sqsubseteq V$ of $\Box$: $\limsup_{U \text{ bounded}} \mathscr{F}_U - \mathscr{F}(\Box) \in \mathscr{D} \cup \{0\}$.

Observe that $\limsup_{U \text{ bounded}} \mathscr{F}_U = \limsup \mathscr{F}$.

Marginal Extension Theorem

Theorem (Marginal Extension Theorem)

There is a smallest set of gambles that satisfies D1–D4, and D5–D6 and that includes $\mathscr{D}$. This natural extension of $\mathscr{D}$ is given by

$$\mathscr{E}_{\mathscr{D}} := \mathscr{G}(\Omega)_{>0} \cup \left\{ g \colon g \geq \limsup \mathscr{I}^{\mathscr{S}} \text{ for some desirable } \square\text{-selection } \mathscr{S} \right\}.$$

Moreover, for any non-terminal situation t and any t-gamble g, $I_{E(t)} g \in \mathscr{E}_{\mathscr{D}}$ iff $g > 0$ or there is some desirable t-selection $\mathscr{S}_t$ such that $g \geq \limsup \mathscr{I}^{\mathscr{S}_t}$.

For any t-gamble f on $E(t)$ and for any non-terminal situation t,

$$\begin{aligned} \underline{P}(f|t) &:= \sup \left\{ \alpha \colon I_{E(t)}(f - \alpha) \in \mathscr{E}_{\mathscr{D}} \right\} \\ &= \sup \left\{ \alpha \colon f - \alpha \geq \limsup \mathscr{I}^{\mathscr{S}} \text{ for some desirable } t\text{-selection } \mathscr{S} \right\}. \end{aligned}$$

Law of Iterated Expectations

Theorem (Law of Iterated Expectations)

Consider any two cuts U and V of a situation t such that U precedes V. Then for all t-gambles f on $E(t)$,

1. $\underline{P}(f|t) = \underline{P}(\underline{P}(f|U)|t)$;
2. $\underline{P}(f|U) = \underline{P}(\underline{P}(f|V)|U)$.