Predictive inference under exchangeability

and the Imprecise Dirichlet Multinomial Model

Gert de Cooman, Jasper De Bock, Márcio Diniz

Ghent University, SYSTeMS

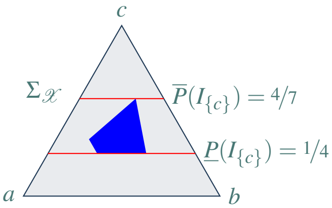
gert.decooman@UGent.be
http://users.UGent.be/~gdcooma
gertekoo.wordpress.com

EBEB 2014
11 March 2014

“It’s probability theory, Jim, but not as we know it”

LOWER PREVISIONS

Lower and upper previsions



Equivalent model

Consider the set $\mathscr{L}(\mathscr{X}) = \mathbb{R}^{\mathscr{X}}$ of all real-valued maps on $\mathscr{X}$. We define two real functionals on $\mathscr{L}(\mathscr{X})$: for all $f\colon \mathscr{X} \to \mathbb{R}$

$$\underline{P}_{\mathscr{M}}(f) = \min\{P_p(f) \colon p \in \mathscr{M}\} \text{ lower prevision/expectation}$$
$$\overline{P}_{\mathscr{M}}(f) = \max\{P_p(f) \colon p \in \mathscr{M}\} \text{ upper prevision/expectation.}$$

Observe that

$$\overline{P}_{\mathscr{M}}(-f) = -\underline{P}_{\mathscr{M}}(f).$$

Basic properties of lower previsions

Definition

We call a real functional $\underline{P}$ on $\mathscr{L}(\mathscr{X})$ a lower prevision if it satisfies the following properties:

for all f and g in $\mathscr{L}(\mathscr{X})$ and all real $\lambda \geq 0$:

1. $\underline{P}(f) \geq \min f$ [boundedness];
2. $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g)$ [super-additivity];
3. $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ [non-negative homogeneity].

Theorem

A real functional $\underline{P}$ is a lower prevision if and only if it is the lower envelope of some credal set $\mathscr{M}$.

Conditioning and lower previsions

Suppose we have two variables X_1 in $\mathscr{X}_1$ and X_2 in $\mathscr{X}_2$.

Consider for instance:

- a joint lower prevision $\underline{P}_{1,2}$ for (X_1, X_2) defined on $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$;
- a conditional lower prevision $\underline{P}_2(\cdot|x_1)$ for X_2 conditional on $X_1 = x_1$, defined on $\mathscr{L}(\mathscr{X}_2)$, for all values $x_1 \in \mathscr{X}_1$.

Coherence

These lower previsions $\underline{P}_{1,2}$ and $\underline{P}_2(\cdot|X_1)$ must satisfy certain (joint) coherence criteria: compare with Bayes's Rule and de Finetti's coherence criteria for precise previsions.

Conditioning and lower previsions

Suppose we have two variables X_1 in $\mathscr{X}_1$ and X_2 in $\mathscr{X}_2$.

Consider for instance:

- a joint lower prevision $\underline{P}_{1,2}$ for (X_1, X_2) defined on $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$;
- a conditional lower prevision $\underline{P}_2(\cdot|x_1)$ for X_2 conditional on $X_1 = x_1$, defined on $\mathscr{L}(\mathscr{X}_2)$, for all values $x_1 \in \mathscr{X}_1$.

Coherence

These lower previsions $\underline{P}_{1,2}$ and $\underline{P}_2(\cdot|X_1)$ must satisfy certain (joint) coherence criteria: compare with Bayes's Rule and de Finetti's coherence criteria for precise previsions.

Complication

A joint lower prevision $\underline{P}_{1,2}$ does not always uniquely determine a conditional $\underline{P}_2(\cdot|X_1)$, we can only impose coherence between them.

Many variables: notation

Suppose we have variables

$$X_i \in \mathscr{X}_i, \quad i \in N$$

For $S \subseteq N$ we denote the S-tuple of variables X_s, $s \in S$ by

$$X_S \in \mathscr{X}_S := \times_{s \in S} \mathscr{X}_s$$

and generic values by $X_S = x_S$.

Many variables: notation

Suppose we have variables

$$X_i \in \mathscr{X}_i, \quad i \in N$$

For $S \subseteq N$ we denote the S-tuple of variables X_s, $s \in S$ by

$$X_S \in \mathscr{X}_S := \times_{s\in S}\mathscr{X}_s$$

and generic values by $X_S = x_S$.

Now consider an input-output pair $I, O \subseteq N$.

Conditional lower previsions

$\underline{P}_O(f(X_O)|x_I) =$ lower prevision of $f(X_O)$ conditional on $X_I = x_I$.

Coherence criterion: Walley (1991)

The conditional lower previsions $\underline{P}_{O_s}(\cdot|X_{I_s})$, $s = 1, \dots, n$ are coherent if and only if:

> for all $f_j \in \mathscr{L}(\mathscr{X}_{O_j \cup I_j})$, all $k \in \{1, \dots, n\}$, all $x_{I_k} \in \mathscr{X}_{I_k}$ and all $g \in \mathscr{L}(\mathscr{X}_{O_k \cup I_k})$, there is some $z_N \in \{x_{I_k}\} \cup \bigcup_{j=1}^{n} \operatorname{supp}_{I_j}(f_j)$ such that:
>
> $$\left[\sum_{s=1}^{n} [f_s - \underline{P}_{O_s}(f_s|x_{I_s})] - [g - \underline{P}_{O_k}(g|x_{I_k})]\right](z_N) \geq 0.$$

where $\operatorname{supp}_I(f) := \{x_I \in \mathscr{X}_I : \mathbb{I}_{\{x_I\}} f \neq 0\}$.

Coherence criterion: Walley (1991)

The conditional lower previsions $\underline{P}_{O_s}(\cdot|X_{I_s})$, $s = 1, \ldots, n$ are coherent if and only if:

> for all $f_j \in \mathscr{L}(\mathscr{X}_{O_j \cup I_j})$, all $k \in \{1, \ldots, n\}$, all $x_{I_k} \in \mathscr{X}_{I_k}$ and all $g \in \mathscr{L}(\mathscr{X}_{O_k \cup I_k})$, there is some $z_N \in \{x_{I_k}\} \cup \bigcup_{j=1}^{n} \mathrm{supp}_{I_j}(f_j)$ such that:
>
> $$\left[\sum_{s=1}^{n} [f_s - \underline{P}_{O_s}(f_s|x_{I_s})] - [g - \underline{P}_{O_k}(g|x_{I_k})]\right](z_N) \geq 0.$$

where $\mathrm{supp}_I(f) := \left\{x_I \in \mathscr{X}_I : \mathbb{I}_{\{x_I\}} f \neq 0\right\}$.

This is quite complicated and cumbersome!

A few papers that try to brave the complications

Available online at www.sciencedirect.com
ScienceDirect
International Journal of Approximate Reasoning
46 (2007) 188–225
ELSEVIER
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

Marginal extension in the theory of coherent lower previsions

Enrique Miranda [a,*], Gert de Cooman [b]

[a] Rey Juan Carlos University, Department of Statistics and Operations Research, C-Tulipán, s/n 28933, Móstoles, Spain
[b] Ghent University, SYSTeMS Research Group, Technologiepark – Zwijnaarde 914, 9052 Zwijnaarde, Belgium

Received 11 July 2006; received in revised form 18 December 2006; accepted 19 December 2006
Available online 3 January 2007

Abstract

We generalise Walley's Marginal Extension Theorem to the case of any finite number of conditional lower previsions. Unlike the procedure of natural extension, our marginal extension always provides the smallest (most conservative) coherent extensions. We show that they can also be calculated as lower envelopes of marginal extensions of conditional linear (precise) previsions. Finally, we use our version of the theorem to study the so-called forward irrelevant product and forward irrelevant natural extension of a number of marginal lower previsions.

Keywords: Imprecise probabilities; Lower previsions; Coherence; Natural extension; Marginal extension; Epistemic irrelevance; Forward irrelevance; Forward irrelevant natural extension; Forward irrelevant product

```
@ARTICLE{miranda2007,
  author = {Miranda, Enrique and De Cooman, Gert},
  title = {Marginal extension in the theory of coherent lower previsions},
  journal = {International Journal of Approximate Reasoning},
  year = 2007,
  volume = 46,
  pages = {188--225}
}
```

A few papers that try to brave the complications

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Journal of Statistical Planning and Inference 138 (2008) 2409–2432

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Weak and strong laws of large numbers for coherent lower previsions☆

Gert De Cooman[a],*, Enrique Miranda[b]

[a] *Systems Research Group, Ghent University, Technologiepark-Zwijnaarde 914, 9052 Zwijnaarde, Belgium*

[b] *Department of Statistics and Operations Research, Rey Juan Carlos University, C-Tulipán, s/n, 28933 Móstoles, Spain*

Received 27 March 2006; received in revised form 13 March 2007; accepted 25 October 2007

Available online 22 November 2007

Abstract

We prove weak and strong laws of large numbers for coherent lower previsions, where the lower prevision of a random variable is given a behavioural interpretation as a subject's supremum acceptable price for buying it. Our laws are a consequence of the rationality criterion of coherence, and they can be proven under assumptions that are surprisingly weak when compared to the standard formulation of the laws in more classical approaches to probability theory.

MSC: 60A99; 60F05; 60F15

Keywords: Imprecise probabilities; Coherent lower previsions; Law of large numbers; Epistemic irrelevance; 2-Monotone capacities

```
@ARTICLE{miranda2007,
  author = {De Cooman, Gert and Miranda, Enrique},
  title = {Weak and strong laws of large numbers for coherent lower previsions},
  journal = {Journal of Statistical Planning and Inference},
  year = 2008,
  volume = 138,
  pages = {2409--2432}
}
```

SETS OF DESIRABLE GAMBLES

Why work with sets of desirable gambles?

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

- is simpler, more intuitive and more elegant
- is more general and expressive than (conditional) lower previsions
- gives a geometrical flavour to probabilistic inference
- shows that probabilistic inference and Bayes' Rule are 'logical' inference
- includes classical propositional logic as another special case
- includes precise probability as one special case
- avoids problems with conditioning on sets of probability zero

First steps: Williams (1977)

Available online at www.sciencedirect.com
ScienceDirect
International Journal of Approximate Reasoning
44 (2007) 366–383

ELSEVIER

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

Notes on conditional previsions ☆

P.M. Williams

Department of Informatics, The University of Sussex, Brighton BN1 9QH, UK

Received 15 December 2005; received in revised form 30 June 2006; accepted 31 July 2006
Available online 25 September 2006

Abstract

The personalist conception of probability is often explicated in terms of betting rates acceptable to an individual. A common approach, that of de Finetti for example, assumes that the individual is willing to take either side of the bet, so that the bet is "fair" from the individual's point of view. This can sometimes be unrealistic, and leads to difficulties in the case of conditional probabilities or previsions. An alternative conception is presented in which it is only assumed that the collection of acceptable bets forms a convex cone, rather than a linear space. This leads to the more general conception of an upper conditional prevision. The main concerns of the paper are with the extension of upper conditional previsions. The main result is that any upper conditional prevision is the upper envelope of a family of additive conditional previsions.

Keywords: Conditional prevision; Imprecise probabilities; Coherence; de Finetti

```
@ARTICLE{williams2007,
   author = {Williams, Peter M.},
   title = {Notes on conditional previsions},
   journal = {International Journal of Approximate Reasoning},
   year = 2007,
   volume = 44,
   pages = {366--383}
}
```

First steps: Walley (2000)

ELSEVIER International Journal of Approximate Reasoning 24 (2000) 125–148

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Towards a unified theory of imprecise probability

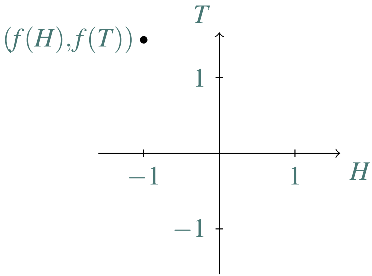
Peter Walley

```
@ARTICLE{walley2000,
   author = {Walley, Peter},
   title = {Towards a unified theory of imprecise probability},
   journal = {International Journal of Approximate Reasoning},
   year = 2000,
   volume = 24,
   pages = {125--148}
}
```

Set of desirable gambles as a belief model

Gambles:

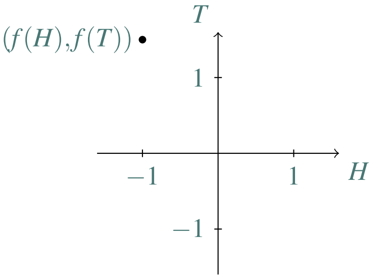
A gamble $f\colon \mathscr{X} \to \mathbb{R}$ is an uncertain reward whose value is $f(X)$.



Set of desirable gambles as a belief model

Gambles:

A gamble $f\colon \mathscr{X} \to \mathbb{R}$ is an uncertain reward whose value is $f(X)$.



Set of desirable gambles:

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

Coherence for a set of desirable gambles

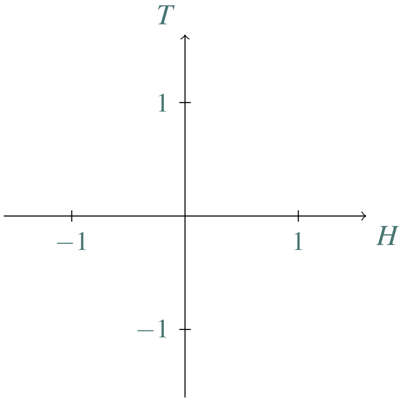
A set of desirable gambles $\mathscr{D}$ is called coherent if:

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

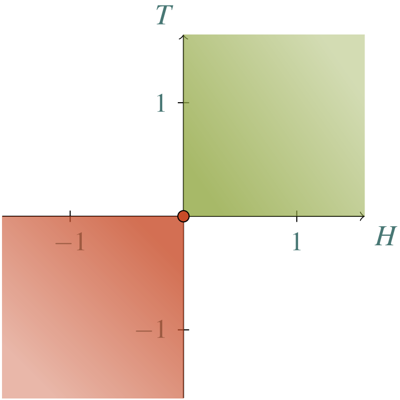
A set of desirable gambles $\mathscr{D}$ is called coherent if:

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

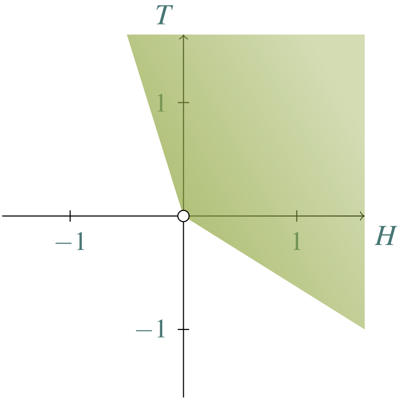
A set of desirable gambles $\mathscr{D}$ is called coherent if:

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:

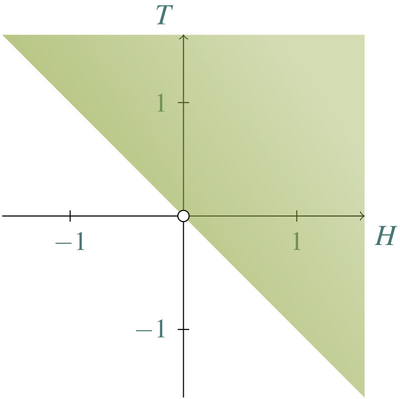
D1. if $f \leq 0$ then $f \notin \mathscr{D}$ [not desiring non-positivity]

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]

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

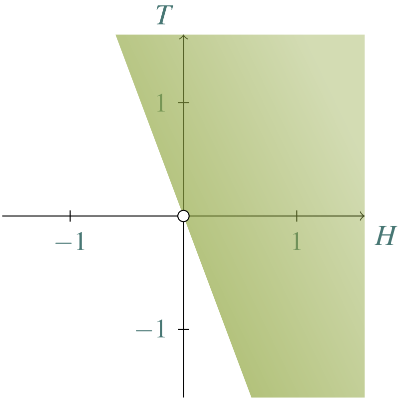


Coherence for a set of desirable gambles

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]

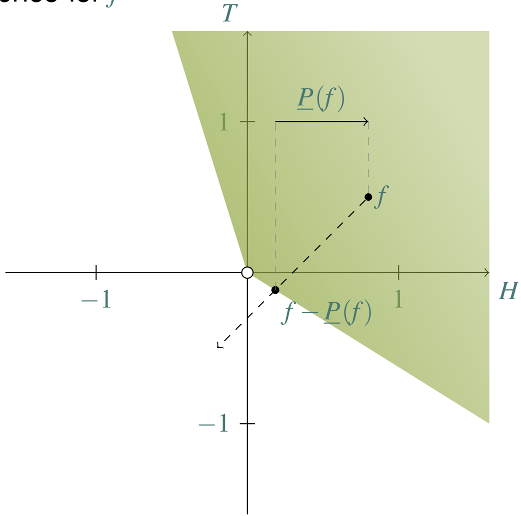
Precise models correspond to the special case that the convex cones $\mathscr{D}$ are actually halfspaces!



Connection with lower previsions

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

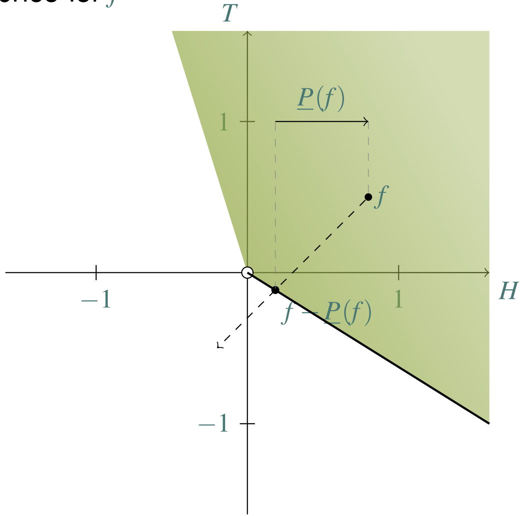
supremum buying price for f



Connection with lower previsions

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

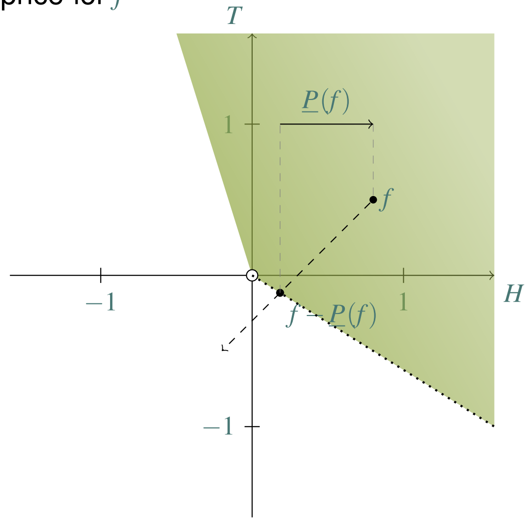
supremum buying price for f



Connection with lower previsions

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

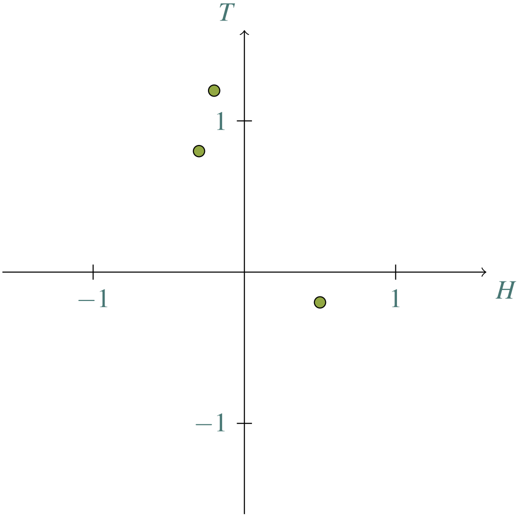
supremum buying price for f



INFERENCE

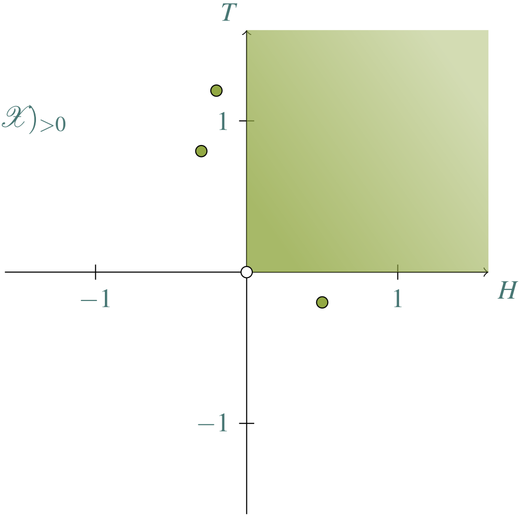
Inference: natural extension

$\mathscr{A}$



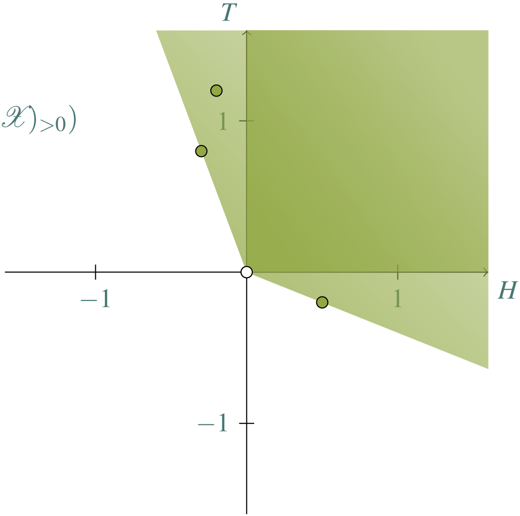
Inference: natural extension

$\mathscr{A} \cup \mathscr{L}(\mathscr{X})_{>0}$



Inference: natural extension

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



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

Inference: marginalisation and conditioning

Let $\mathscr{D}_N$ be a coherent set of desirable gambles on $\mathscr{X}_N$.

For any subset $I \subseteq N$, we have the $\mathscr{X}_I$-marginals:

$$\mathscr{D}_I = \operatorname{marg}_I(\mathscr{D}_N) := \mathscr{D}_N \cap \mathscr{L}(\mathscr{X}_I),$$

so

$$f(X_I) \in \mathscr{D}_I \Leftrightarrow f(X_I) \in \mathscr{D}_N.$$

How to condition a coherent set $\mathscr{D}_N$ on the observation that $X_I = x_I$?

The updated set of desirable gambles $\mathscr{D}_N \rfloor x_I \subseteq \mathscr{L}(\mathscr{X}_{N\setminus I})$ on $\mathscr{X}_{N\setminus I}$ is:

$$g \in \mathscr{D}_N \rfloor x_I \Leftrightarrow \mathbb{I}_{\{x_I\}} g \in \mathscr{D}_N.$$

Works for all conditioning events: no problem with conditioning on sets of probability zero!

Conditional lower previsions

Just like in the unconditional case, we can use a coherent set of desirable gambles $\mathscr{D}_N$ to derive conditional lower previsions.

Consider disjoint subsets I and O of N:

$$\begin{aligned}\underline{P}_O(g|x_I) &:= \sup\left\{\mu \in \mathbb{R} \colon \mathbb{I}_{\{x_I\}}[g-\mu] \in \mathscr{D}_N\right\} \\ &= \sup\{\mu \in \mathbb{R} \colon g-\mu \in \mathscr{D}_N\rfloor x_I\} \text{ for all } g \in \mathscr{L}(\mathscr{X}_O)\end{aligned}$$

is the lower prevision of g, conditional on $X_I = x_I$.

$\underline{P}_O(g|X_I)$ is the gamble on $\mathscr{X}_I$ that assumes the value $\underline{P}_O(g|x_I)$ in $x_I \in \mathscr{X}_I$.

Coherent conditional lower previsions

Consider m couples of disjoint subsets I_s and O_s of N, and corresponding conditional lower previsions $\underline{P}_{O_s}(\cdot|X_{I_s})$ for $s = 1, \dots, m$.

Theorem (Williams, 1977)

These conditional lower previsions are (jointly) coherent if and only if there is some coherent set of desirable gambles $\mathscr{D}_N$ that produces them, in the sense that for all $s = 1, \dots, m$:

$$\underline{P}_{O_s}(g|x_{I_s}) := \sup\left\{\mu \in \mathbb{R} \colon \mathbb{I}_{\{x_{I_s}\}}[g - \mu] \in \mathscr{D}_N\right\}$$

for all $g \in \mathscr{L}(\mathscr{X}_{O_s})$ and all $x_{I_s} \in \mathscr{X}_{I_s}$.

A few papers that avoid the complications

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

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{miranda2007,
   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}
}
```

A few papers that avoid the complications

Journal of Artificial Intelligence Research 45 (2012) 601-640 Submitted 8/12; published 12/12

Irrelevant and Independent Natural Extension for Sets of Desirable Gambles

Gert de Cooman GERT.DECOOMAN@UGENT.BE
Ghent University, SYSTeMS Research Group
Technologiepark 914, 9052 Zwijnaarde, Belgium

Enrique Miranda MIRANDAENRIQUE@UNIOVI.ES
University of Oviedo, Dept. of Statistics and O.R.
C-Calvo Sotelo, s/n, Oviedo 33007, Spain

Abstract

The results in this paper add useful tools to the theory of sets of desirable gambles, a growing toolbox for reasoning with partial probability assessments. We investigate how to combine a number of marginal coherent sets of desirable gambles into a joint set using the properties of epistemic irrelevance and independence. We provide formulas for the smallest such joint, called their independent natural extension, and study its main properties. The independent natural extension of maximal coherent sets of desirable gambles allows us to define the strong product of sets of desirable gambles. Finally, we explore an easy way to generalise these results to also apply for the conditional versions of epistemic irrelevance and independence. Having such a set of tools that are easily implemented in computer programs is clearly beneficial to fields, like AI, with a clear interest in coherent reasoning under uncertainty using general and robust uncertainty models that require no full specification.

```
@ARTICLE{miranda2007,
   author = {De Cooman, Gert and Miranda, Enrique},
   title = {Irrelevant and independent natural extension for sets of desirable gambles},
   journal = {Journal of Artificial Intelligence Research},
   year = 2012,
   volume = 45,
   pages = {601--640}
}
```

PERMUTATION SYMMETRY

Symmetry group

Consider a variable X assuming values in a finite set $\mathscr{X}$ and a finite group $\mathscr{P}$ of permutations π of $\mathscr{X}$

Modelling that there is a symmetry $\mathscr{P}$ behind X:

if you believe that inferences about X will be invariant under any permutation $\pi \in \mathscr{P}$.

Consider any permutation $\pi \in \mathscr{P}$ and any gamble f on $\mathscr{X}$:

$$\pi^t f := f \circ \pi, \text{ meaning that } (\pi^t f)(x) = f(\pi(x)) \text{ for all } x \in \mathscr{X}.$$

π^t is a linear transformation of the vector space $\mathscr{L}(\mathscr{X})$.

$$\mathscr{P}^t := \{\pi^t : \pi \in \mathscr{P}\}$$

is a finite group of linear transformations of the vector space $\mathscr{L}(\mathscr{X})$.

How to model permutation symmetry?

Permutation symmetry:

You are indifferent between any gamble f on $\mathscr{X}$ and its permutation $\pi^t f$.

$$f \approx \pi^t f \Leftrightarrow f - \pi^t f \approx 0.$$

This leads to a linear space $\mathscr{I}_{\mathscr{P}}$ of indifferent gambles:

$$\mathscr{I}_{\mathscr{P}} := \operatorname{span}(\{f - \pi^t f : f \in \mathscr{L}(\mathscr{X}) \text{ and } \pi \in \mathscr{P}\}).$$

How to model permutation symmetry?

Accept & Reject Statement-Based Uncertainty Models

Erik Quaeghebeur, Gert de Cooman, and Filip Hermans*
SYSTeMS Research Group, Ghent University, Gent, Belgium.

Abstract. We develop a framework for modelling and reasoning with uncertainty based on accept and reject statements about gambles. It generalises the frameworks found in the literature based on statements of acceptability, desirability, or favourability and clarifies their relative position. Next to the statement-based formulation, we also provide a translation in terms of preference relations, discuss—as a bridge to existing frameworks—a number of simplified variants, and show the relationship with prevision-based uncertainty models. We furthermore provide an application to modelling symmetry judgements.

Keywords: acceptability, indifference, desirability, favourability, preference, prevision

How to model permutation symmetry?

Permutation symmetry:

You are indifferent between any gamble f on $\mathscr{X}$ and its permutation $\pi^t f$.

$$f \approx \pi^t f \Leftrightarrow f - \pi^t f \approx 0.$$

This leads to a linear space $\mathscr{I}_{\mathscr{P}}$ of indifferent gambles:

$$\mathscr{I}_{\mathscr{P}} := \operatorname{span}(\{f - \pi^t f : f \in \mathscr{L}(\mathscr{X}) \text{ and } \pi \in \mathscr{P}\}).$$

A set of desirable gambles $\mathscr{D}$ is strongly $\mathscr{P}$-invariant if

$$\mathscr{D} + \mathscr{I}_{\mathscr{P}} \subseteq \mathscr{D}$$

or actually:

$$\boxed{\mathscr{D} + \mathscr{I}_{\mathscr{P}} = \mathscr{D}.}$$

Let's see what this means:

Consider the linear subspace of all permutation invariant gambles:

$$\mathscr{L}_{\mathscr{P}}(\mathscr{X}) := \left\{ f \in \mathscr{L}(\mathscr{X}) \colon (\forall \pi \in \mathscr{P}) f = \pi^t f \right\}$$

Because $\mathscr{P}^t$ is a finite group, this linear space can be seen as the range of the following special linear transformation:

$$\operatorname{proj}_{\mathscr{P}} \colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) \text{ with } \operatorname{proj}_{\mathscr{P}}(f) := \frac{1}{|\mathscr{P}|} \sum_{\pi \in \mathscr{P}} \pi^t f.$$

Theorem

1. $\pi^t \circ \operatorname{proj}_{\mathscr{P}} = \operatorname{proj}_{\mathscr{P}} = \operatorname{proj}_{\mathscr{P}} \circ \pi^t$
2. $\operatorname{proj}_{\mathscr{P}} \circ \operatorname{proj}_{\mathscr{P}} = \operatorname{proj}_{\mathscr{P}}$
3. $\operatorname{rng}(\operatorname{proj}_{\mathscr{P}}) = \mathscr{L}_{\mathscr{P}}(\mathscr{X})$ *and*
4. $\ker(\operatorname{proj}_{\mathscr{P}}) = \mathscr{I}_{\mathscr{P}}$.

Symmetry representation theorem

Consider any gamble f on $\mathscr{X}$:

$$f = \underbrace{f - \mathrm{proj}_{\mathscr{P}}(f)}_{\in \mathscr{I}_{\mathscr{P}}} + \underbrace{\mathrm{proj}_{\mathscr{P}}(f)}_{\in \mathscr{L}_{\mathscr{P}}(\mathscr{X})}$$

Theorem (Symmetry Representation Theorem)

Let $\mathscr{D}$ be strongly $\mathscr{P}$-invariant, then:

$$f \in \mathscr{D} \Leftrightarrow \mathrm{proj}_{\mathscr{P}}(f) \in \mathscr{D}.$$

So $\mathscr{D}$ has a lower-dimensional representation $\mathrm{proj}_{\mathscr{P}}(\mathscr{D}) \subseteq \mathscr{L}_{\mathscr{P}}(\mathscr{X})$.

Permutation invariant gambles

Consider the permutation invariant atoms:

$$[x] := \{\pi(x) \colon \pi \in \mathscr{P}\}$$

which constitute a partition of $\mathscr{X}$:

$$\mathscr{A}_{\mathscr{P}} := \{[x] \colon x \in \mathscr{X}\}.$$

Then the permutation invariant gambles $f \in \mathscr{L}_{\mathscr{P}}(\mathscr{X})$

- are constant on these invariant atoms $[x]$; and
- can therefore be seen as gambles on these atoms.

$$\text{“}f \in \mathscr{L}(\mathscr{A}_{\mathscr{P}})\text{”}.$$

FINITE EXCHANGEABILITY

Exchangeability for sets of desirable gambles

International Journal of Approximate Reasoning 53 (2012) 363–395

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Exchangeability and sets of desirable gambles

Gert de Cooman, Erik Quaeghebeur *

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

A R T I C L E   I N F O

Article history:
Available online 13 December 2010

Keywords:
Sets of desirable gambles
Exchangeability
Representation
Natural extension
Updating
Extending an exchangeable sequence

A B S T R A C T

Sets of desirable gambles constitute a quite general type of uncertainty model with an interesting geometrical interpretation. We give a general discussion of such models and their rationality criteria. We study exchangeability assessments for them, and prove counterparts of de Finetti's Finite and Infinite Representation Theorems. We show that the finite representation in terms of count vectors has a very nice geometrical interpretation, and that the representation in terms of frequency vectors is tied up with multivariate Bernstein (basis) polynomials. We also lay bare the relationships between the representations of updated exchangeable models, and discuss conservative inference (natural extension) under exchangeability and the extension of exchangeable sequences.


```
@ARTICLE{cooman2012,
  author = {De Cooman, Gert and Quaeghebeur, Erik},
  title = {Exchangeability and sets of desirable gambles},
  journal = {International Journal of Approximate Reasoning},
  year = 2012,
  vol = 53,
  pages = {363--395}
}
```

Permutations and count vectors

Consider any permutation π of the set of indices $\{1,2,\ldots,n\}$.

For any $x=(x_1,x_2,\ldots,x_n)$ in $\mathscr{X}^n$, we let

$$\pi x := (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}).$$

For any $x \in \mathscr{X}^n$, consider the corresponding count vector $T(x)$, where for all $z \in \mathscr{X}$:

$$T_z(x) := |\{k \in \{1,\ldots,n\} : x_k = z\}|.$$

Example

For $\mathscr{X} = \{a,b\}$ and $x = (a,a,b,b,a,b,b,a,a,a,b,b,b)$, we have

$$T_a(x) = 6 \text{ and } T_b(x) = 7.$$

Let $m = T(x)$ and consider the permutation invariant atom

$$[m] := \{y \in \mathscr{X}^n : T(y) = m\}.$$

This atom has how many elements?

$$\binom{n}{m} = \frac{n!}{\prod_{x \in \mathscr{X}} m_x!}$$

Let $\mathrm{HypGeo}^n(\cdot|m)$ be the expectation operator associated with the uniform distribution on $[m]$:

$$\mathrm{HypGeo}^n(f|m) := \binom{n}{m}^{-1} \sum_{x \in [m]} f(x) \text{ for all } f : \mathscr{X}^n \to \mathbb{R}$$

Interestingly:

$$\mathrm{proj}_{\mathscr{P}}(f)(x) = \mathrm{HypGeo}^n(f|m) \text{ for all } x \in [m]$$

can be seen as a gamble on the composition m of an urn with n balls.

COUNTABLE EXCHANGEABILITY

Bruno de Finetti's exchangeability result

Infinite Representation Theorem:

The sequence $X_1, \dots, X_n, \dots$ of random variables in the finite set $\mathscr{X}$ is exchangeable iff there is a (unique) coherent prevision H on the linear space $\mathscr{V}(\Sigma_{\mathscr{X}})$ of all polynomials on $\Sigma_{\mathscr{X}}$ such that for all $n \in \mathbb{N}$ and all $f\colon \mathscr{X}^n \to \mathbb{R}$:

$$E(f) = H\left(\sum_{m \in \mathscr{N}^n} \mathrm{HypGeo}^n(f|m) B_m\right).$$

Observe that

$$B_m(\theta) = \mathrm{MultiNom}^n([m]|\theta) = \binom{n}{m} \prod_{x \in \mathscr{X}} \theta_x^{m_x}$$

$$\sum_{m \in \mathscr{N}^n} \mathrm{HypGeo}^n(f|m) B_m(\theta) = \mathrm{MultiNom}^n(f|\theta)$$

Representation theorem for sets of desirable gambles

Representation Theorem

A sequence $\mathscr{D}^1, \ldots, \mathscr{D}^n, \ldots$ of coherent sets of desirable gambles is exchangeable iff there is some (unique) Bernstein coherent $\mathscr{H} \subseteq \mathscr{V}(\Sigma_{\mathscr{X}})$ such that:

$$f \in \mathscr{D}^n \Leftrightarrow \mathrm{MultiNom}^n(f|\cdot) \in \mathscr{H} \quad \text{for all } n \in \mathbb{N} \text{ and } f \in \mathscr{L}(\mathscr{X}^n).$$

A set $\mathscr{H}$ of polynomials on $\Sigma_{\mathscr{X}}$ is Bernstein coherent if:

B1. if p has some non-positive Bernstein expansion then $p \notin \mathscr{H}$

B2. if p has some positive Bernstein expansion then $p \in \mathscr{H}$

B3. if $p_1 \in \mathscr{H}$ and $p_2 \in \mathscr{H}$ then $p_1 + p_2 \in \mathscr{H}$

B4. if $p \in \mathscr{H}$ then $\lambda p \in \mathscr{H}$ for all positive real numbers λ.

Suppose we observe the first n variables, with count vector $m = T(x)$:

$$(X_1, \ldots, X_n) = (x_1, \ldots, x_n) = x.$$

Then the remaining variables

$$X_{n+1}, \ldots, X_{n+k}, \ldots$$

are still exchangeable, with representation $\mathscr{H}\rfloor x = \mathscr{H}\rfloor m$ given by:

$$p \in \mathscr{H}\rfloor m \Leftrightarrow B_m p \in \mathscr{H}.$$

Suppose we observe the first n variables, with count vector $m = T(x)$:

$$(X_1, \ldots, X_n) = (x_1, \ldots, x_n) = x.$$

Then the remaining variables

$$X_{n+1}, \ldots, X_{n+k}, \ldots$$

are still exchangeable, with representation $\mathscr{H} \rfloor x = \mathscr{H} \rfloor m$ given by:

$$p \in \mathscr{H} \rfloor m \Leftrightarrow B_m p \in \mathscr{H}.$$

Conclusion:

A Bernstein coherent set of polynomials $\mathscr{H}$ completely characterises all predictive inferences about an exchangeable sequence.

PREDICTIVE INFERENCE SYSTEMS

Predictive inference systems

Formal definition:

A predictive inference system is a map Ψ that associates with every finite set of categories $\mathscr{X}$ a Bernstein coherent set of polynomials on $\Sigma_{\mathscr{X}}$:

$$\Psi(\mathscr{X}) = \mathscr{H}_{\mathscr{X}}.$$

Basic idea:

Once the set of possible observations $\mathscr{X}$ is determined, then all predictive inferences about successive observations $X_1, \ldots, X_n, \ldots$ in $\mathscr{X}$ are completely fixed by $\Psi(\mathscr{X}) = \mathscr{H}_{\mathscr{X}}$.

Inference principles

Even if (when) you don't like this idea, you might want to concede the following:

Using inference principles to constrain Ψ:

We can use general inference principles to impose conditions on Ψ, or in other words to constrain:

- the values $\mathscr{H}_{\mathscr{X}}$ can assume for different $\mathscr{X}$
- the relation between $\mathscr{H}_{\mathscr{X}}$ and $\mathscr{H}_{\mathscr{Y}}$ for different $\mathscr{X}$ and $\mathscr{Y}$

Inference principles

Even if (when) you don't like this idea, you might want to concede the following:

Using inference principles to constrain Ψ:

We can use general inference principles to impose conditions on Ψ, or in other words to constrain:

- the values $\mathscr{H}_{\mathscr{X}}$ can assume for different $\mathscr{X}$
- the relation between $\mathscr{H}_{\mathscr{X}}$ and $\mathscr{H}_{\mathscr{Y}}$ for different $\mathscr{X}$ and $\mathscr{Y}$

Taken to (what might be called) extremes (Carnap, Walley, ...):

Impose so many constraints (principles) that you end up with a single Ψ, or a parametrised family of them, e.g.:

the λ system, the IDM family

A few examples

Renaming Invariance

Inferences should not be influenced by what names we give to the categories.

Pooling Invariance

For gambles that do not differentiate between pooled categories, it should not matter whether we consider predictive inferences for the set of original categories $\mathscr{X}$, or for the set of pooled categories $\mathscr{Y}$.

Specificity

If, after making a number of observations in $\mathscr{X}$, we decide that in the future we are only going to consider outcomes in a subset $\mathscr{Y}$ of $\mathscr{X}$, we can discard from the past observations those outcomes not in $\mathscr{Y}$.

Look how nice!

Renaming Invariance

For any onto and one-to-one $\pi : \mathscr{X} \to \mathscr{Y}$

$$B_{C_\pi(m)} p \in \mathscr{H}_\mathscr{Y} \Leftrightarrow B_m(p \circ C_\pi) \in \mathscr{H}_\mathscr{X} \text{ for all } p \in \mathscr{V}(\Sigma_\mathscr{Y}) \text{ and } m \in \mathscr{N}_\mathscr{X}$$

Pooling Invariance

For any onto $\rho : \mathscr{X} \to \mathscr{Y}$

$$B_{C_\rho(m)} p \in \mathscr{H}_\mathscr{Y} \Leftrightarrow B_m(p \circ C_\rho) \in \mathscr{H}_\mathscr{X} \text{ for all } p \in \mathscr{V}(\Sigma_\mathscr{Y}) \text{ and } m \in \mathscr{N}_\mathscr{X}$$

Specificity

For any $\mathscr{Y} \subseteq \mathscr{X}$

$$B_{m|_\mathscr{Y}} p \in \mathscr{H}_\mathscr{Y} \Leftrightarrow B_m(p \circ \cdot|_\mathscr{Y}) \in \mathscr{H}_\mathscr{X} \text{ for all } p \in \mathscr{V}(\Sigma_\mathscr{Y}) \text{ and } m \in \mathscr{N}_\mathscr{X}$$

EXAMPLES

Imprecise Dirichlet Multinomial Model

Let, with $s > 0$:

$$\Delta^s_{\mathscr{X}} := \left\{ \alpha \in \mathbb{R}^{\mathscr{X}}_{>0} \colon \sum_{x \in \mathscr{X}} \alpha_x < s \right\}$$

and

$$\mathscr{H}^s_{\mathscr{X}} := \{ p \in \mathscr{V}(\Sigma_{\mathscr{X}}) \colon (\forall \alpha \in \Delta^s_{\mathscr{X}}) \operatorname{Diri}(p|\alpha) > 0 \}$$

This inference system is pooling and renaming invariant, and specific.

For any observed count vector $m \in \mathscr{N}^n$:

$$\underline{P}_s(\{x\}|m) = \frac{m_x}{n+s} \text{ and } \overline{P}_s(\{x\}|m) = \frac{m_x + s}{n+s}$$

Haldane system

Let:

$$\begin{aligned}\mathscr{H}_{\mathscr{X}}^{H} &:= \bigcup_{s>0} \{p \in \mathscr{V}(\Sigma_{\mathscr{X}}) \colon (\forall \alpha \in \Delta_{\mathscr{X}}^{s})\,\mathrm{Diri}(p|\alpha) > 0\} \\ &= \bigcup_{s>0} \mathscr{H}_{\mathscr{X}}^{s}\end{aligned}$$

This inference system is pooling and renaming invariant, and specific.

For any observed count vector $m \in \mathscr{N}^{n}$:

$$\underline{P}_s(\{x\}|m) = \overline{P}_s(\{x\}|m) = \frac{m_x}{n}$$