Probabilistic Logic and Probabilistic Networks

Rolf Haenni

Reasoning under **UN**certainty Group
Institute of Computer Science and Applied Mathematics
University of Berne, Switzerland

Progic'07
3rd Workshop on Combining Probability and Logic
University of Kent
September 5–7, 2007

Outline

Part I: Probabilistic Logic

Part II: Probabilistic Argumentation

Part III: Probabilistic Networks

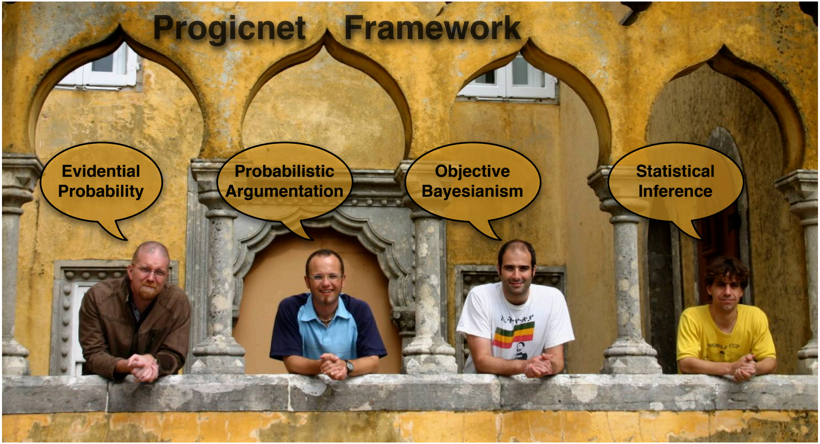
Progicnet Academic Network

Gregory Wheeler	**Rolf Haenni**	**Jon Williamson**	**Jan-Willem Romeijn**
New Univ. of Lisbon	Univ. of Bern	Univ. of Kent	Univ. of Groningen
Portugal	Switzerland	United Kingdom	The Netherlands

Progicnet Academic Network

Gregory Wheeler	**Rolf Haenni**	**Jon Williamson**	**Jan-Willem Romeijn**
New Univ. of Lisbon	Univ. of Bern	Univ. of Kent	Univ. of Groningen
Portugal	Switzerland	United Kingdom	The Netherlands

Progicnet Academic Network



Gregory Wheeler	**Rolf Haenni**	**Jon Williamson**	**Jan-Willem Romeijn**
New Univ. of Lisbon	Univ. of Bern	Univ. of Kent	Univ. of Groningen
Portugal	Switzerland	United Kingdom	The Netherlands

Project Overview

- Sponsored by "The Leverhulme Trust"
- Two-year period (April'06–March'08)
- Interdisciplinary project
  - Gregory & Rolf: Computer Science
  - Jon: Philosophy
  - Jan-Willem: Psychology & Philosophy
- Project goals
  - Promote and advance the research on probabilistic logic
  - Connect different logical and probabilistic inferential systems
  - Apply probabilistic networks to probabilistic logic
  - Exchange ideas, experience, knowledge
  - Common publications
- Homepage
  - www.kent.ac.uk/secl/philosophy/jw/2006/progicnet.htm

Activities

- Regular meetings
  - Canterbury, April'06
  - Lisbon, September'06
  - Leukerbad/Berne, January'07
  - Amsterdam, May'07
  - Canterbury, September'07
  - Granada, February'08
- Common talks
  - FotFS'07, 6th Conf. on Foundations of the Formal Sciences
  - FEW'07, 4th Annual Formal Epistemology Workshop
  - Workshop on Methodological Problems of the Social Sciences
  - Progic'07, 3rd Workshop on Combining Probability and Logic
- 3rd Progic workshop
- Special issue of the Journal of Applied Logic

Outline

Part I: Probabilistic Logic

1. The Potential of Probabilistic Logic
2. Standard Probabilistic Semantics

Outline

Part I: Probabilistic Logic

1. The Potential of Probabilistic Logic
2. Standard Probabilistic Semantics

Progic Framework

- Classical logical inference concerns truth value assignments, while inference with probabilistic logic concerns probability assignments
- Inference in classical logic: $\varphi_1, \ldots, \varphi_n \models \psi$?
  - premises φ_i
  - conclusion ψ
  - decide yes/no
- Inference in probabilistic logic: $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^{Y}$?
  - probability sets $X_i \subseteq [0, 1]$
  - find $Y \subseteq [0, 1]$

$\Rightarrow$ **Progic framework**

Scope & Motivation

- Potential application areas of the Progic framework are
  - formal epistemology
  - mathematical statistics
  - philosophy of science
  - artificial intelligence
  - bioinformatics
  - linguistics
- But probabilistic logics are not widely used, because they seem to be
  - disparate
  - hard to understand
  - computationally complex
  - not well established

Progicnet Strategy

- Demonstrate that several probabilistic logics can be brought under the unifying umbrella of the Progic framework
  - classical and Bayesian statistics (Jan-Willem)
  - evidential probability (Gregory)
  - objective Bayesianism (Jon)
  - probabilistic argumentation (Rolf)
- Thus, the strategy is to show that each of these paradigms
  a) is representable as
  b) provides semantics for

  questions of the general form $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^Y$?
- To better handle the computational complexity, the strategy is to link the Progic framework to probabilistic (credal) networks

Outline

Part I: Probabilistic Logic

1. The Potential of Probabilistic Logic

2. Standard Probabilistic Semantics

General Idea

- Each premise $\varphi_i^{X_i}$ in $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^{Y?}$ is interpreted as a constraint $P(\varphi_i) \in X_i$ for the unknown prob. measure $P \in \mathbb{P}$
- The combined constraints of the premises may be

under-determined $\Rightarrow$ non-empty set $\mathbb{P}_* \subseteq \mathbb{P}$ of probability measures
just right $\Rightarrow$ single probability measure $\mathbb{P}_* = \{P\}$
over-determined $\Rightarrow$ $\mathbb{P}_* = \emptyset$, i.e. something is wrong

$$\varphi_1^{X_1}, \varphi_2^{X_2} \models \psi^{Y?}$$

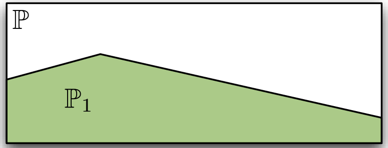
- General (under-determined) case: $Y = \{P(\psi) : P \in \mathbb{P}_*\}$
- Note that even if all sets X_i are singletons, i.e. $X_i = \{x_i\}$, we may still get non-singletons for Y

General Idea

- Each premise $\varphi_i^{X_i}$ in $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^{Y?}$ is interpreted as a constraint $P(\varphi_i) \in X_i$ for the unknown prob. measure $P \in \mathbb{P}$
- The combined constraints of the premises may be

under-determined $\Rightarrow$ non-empty set $\mathbb{P}_* \subseteq \mathbb{P}$ of probability measures
just right $\Rightarrow$ single probability measure $\mathbb{P}_* = \{P\}$
over-determined $\Rightarrow$ $\mathbb{P}_* = \emptyset$, i.e. something is wrong

$$\varphi_1^{X_1}, \varphi_2^{X_2} \models \psi^{Y?}$$



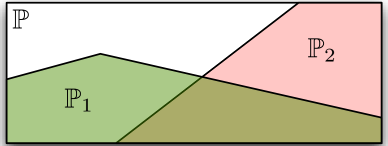
- General (under-determined) case: $Y = \{P(\psi) : P \in \mathbb{P}_*\}$
- Note that even if all sets X_i are singletons, i.e. $X_i = \{x_i\}$, we may still get non-singletons for Y

General Idea

- Each premise $\varphi_i^{X_i}$ in $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^{Y?}$ is interpreted as a constraint $P(\varphi_i) \in X_i$ for the unknown prob. measure $P \in \mathbb{P}$
- The combined constraints of the premises may be

  under-determined $\Rightarrow$ non-empty set $\mathbb{P}_* \subseteq \mathbb{P}$ of probability measures
  just right $\Rightarrow$ single probability measure $\mathbb{P}_* = \{P\}$
  over-determined $\Rightarrow$ $\mathbb{P}_* = \emptyset$, i.e. something is wrong

$$\varphi_1^{X_1}, \varphi_2^{X_2} \models \psi^{Y?}$$



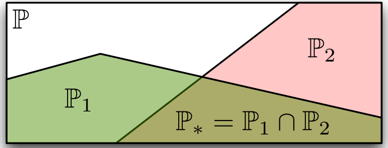
- General (under-determined) case: $Y = \{P(\psi) : P \in \mathbb{P}_*\}$
- Note that even if all sets X_i are singletons, i.e. $X_i = \{x_i\}$, we may still get non-singletons for Y

General Idea

- Each premise $\varphi_i^{X_i}$ in $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^{Y?}$ is interpreted as a constraint $P(\varphi_i) \in X_i$ for the unknown prob. measure $P \in \mathbb{P}$
- The combined constraints of the premises may be

  under-determined $\Rightarrow$ non-empty set $\mathbb{P}_* \subseteq \mathbb{P}$ of probability measures
  just right $\Rightarrow$ single probability measure $\mathbb{P}_* = \{P\}$
  over-determined $\Rightarrow$ $\mathbb{P}_* = \emptyset$, i.e. something is wrong

$$\varphi_1^{X_1}, \varphi_2^{X_2} \models \psi^{Y?}$$

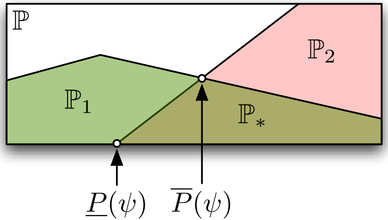


- General (under-determined) case: $Y = \{P(\psi) : P \in \mathbb{P}_*\}$
- Note that even if all sets X_i are singletons, i.e. $X_i = \{x_i\}$, we may still get non-singletons for Y

Probability Intervals

- If all probability sets X_i are (functionally unrelated) intervals, i.e. sub-intervals of $[0,1]$, then
  - all sets $\mathbb{P}_i$ are convex
  - $\mathbb{P}_*$ is also convex
  - Y is also an interval, i.e. $Y = [\underline{P}(\psi), \overline{P}(\psi)]$, where $\underline{P}$ and $\overline{P}$ are vertices of $\mathbb{P}$

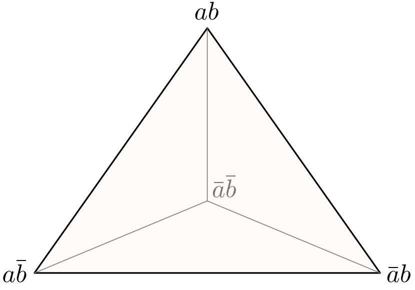
$$\varphi_1^{X_1}, \varphi_2^{X_2} \models \psi^{Y?}$$



- Under the standard semantics, inference means to solve very large linear programming problems

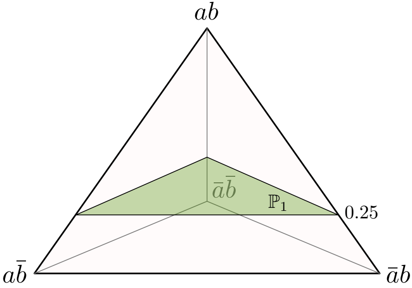
Example 1

- For the premises $(a \wedge b)^{\{0.25\}}, (a \vee \neg b)^{\{1\}}$ we get
  - $Y = [0.25, 1]$, for $\psi = a$
  - $Y = [0.25, 0.25] = \{0.25\}$, for $\psi = b$
  - $Y = [0, 1]$, for $\psi = c$
  - etc.



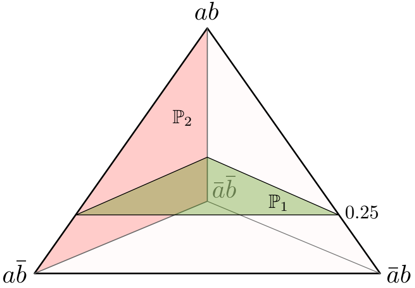
Example 1

- For the premises $(a \wedge b)^{\{0.25\}}, (a \vee \neg b)^{\{1\}}$ we get
  - $Y = [0.25, 1]$, for $\psi = a$
  - $Y = [0.25, 0.25] = \{0.25\}$, for $\psi = b$
  - $Y = [0, 1]$, for $\psi = c$
  - etc.



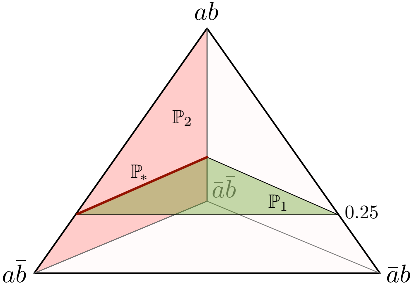
Example 1

- For the premises $(a \wedge b)^{\{0.25\}}, (a \vee \neg b)^{\{1\}}$ we get
  - $Y = [0.25, 1]$, for $\psi = a$
  - $Y = [0.25, 0.25] = \{0.25\}$, for $\psi = b$
  - $Y = [0, 1]$, for $\psi = c$
  - etc.



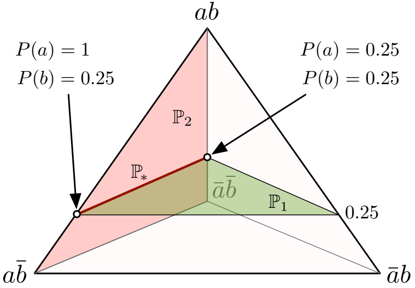
Example 1

- For the premises $(a \wedge b)^{\{0.25\}}, (a \vee \neg b)^{\{1\}}$ we get
  - $Y = [0.25, 1]$, for $\psi = a$
  - $Y = [0.25, 0.25] = \{0.25\}$, for $\psi = b$
  - $Y = [0, 1]$, for $\psi = c$
  - etc.



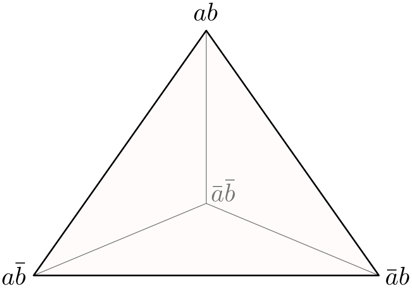
Example 1

- For the premises $(a \wedge b)^{\{0.25\}}, (a \vee \neg b)^{\{1\}}$ we get
  - $Y = [0.25, 1]$, for $\psi = a$
  - $Y = [0.25, 0.25] = \{0.25\}$, for $\psi = b$
  - $Y = [0, 1]$, for $\psi = c$
  - etc.



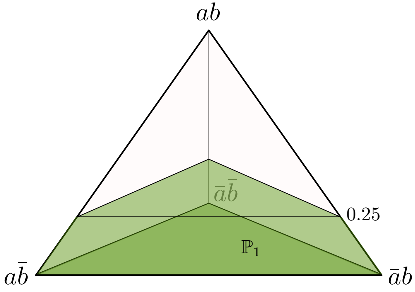
Example 2

- For the premises $(a \wedge b)^{[0,0.25]}, (a \vee \neg b)^{\{1\}}$ we get
  - $Y = [0, 1]$, for $\psi = a$
  - $Y = [0, 0.25]$, for $\psi = b$
  - $Y = [0, 1]$, for $\psi = c$
  - etc.



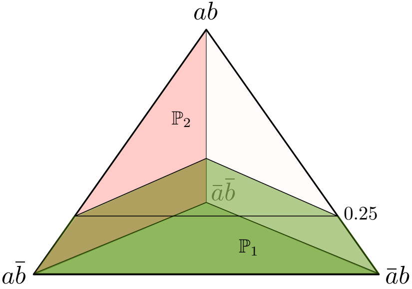
Example 2

- For the premises $(a \wedge b)^{[0,0.25]}, (a \vee \neg b)^{\{1\}}$ we get
  - $Y = [0,1]$, for $\psi = a$
  - $Y = [0,0.25]$, for $\psi = b$
  - $Y = [0,1]$, for $\psi = c$
  - etc.



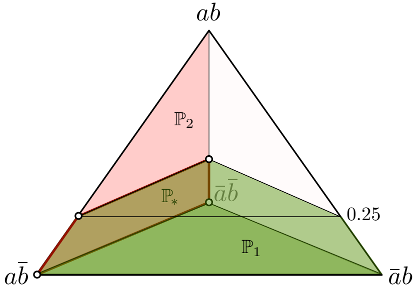
Example 2

- For the premises $(a \wedge b)^{[0,0.25]}, (a \vee \neg b)^{\{1\}}$ we get
  - $Y = [0, 1]$, for $\psi = a$
  - $Y = [0, 0.25]$, for $\psi = b$
  - $Y = [0, 1]$, for $\psi = c$
  - etc.



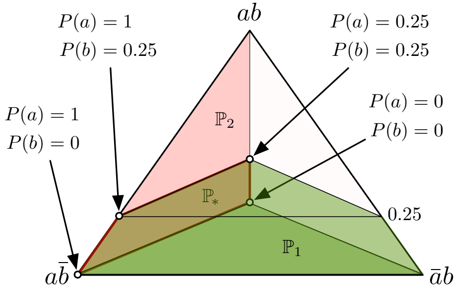
Example 2

- For the premises $(a \wedge b)^{[0,0.25]}, (a \vee \neg b)^{\{1\}}$ we get
  - $Y = [0, 1]$, for $\psi = a$
  - $Y = [0, 0.25]$, for $\psi = b$
  - $Y = [0, 1]$, for $\psi = c$
  - etc.



Example 2

- For the premises $(a \wedge b)^{[0,0.25]}, (a \vee \neg b)^{\{1\}}$ we get
  - $Y = [0, 1]$, for $\psi = a$
  - $Y = [0, 0.25]$, for $\psi = b$
  - $Y = [0, 1]$, for $\psi = c$
  - etc.



Outline

Part II: Probabilistic Argumentation

Outline

Part II: Probabilistic Argumentation

3. Background of Probabilistic Argumentation

4. Connections to Progic Framework

Motivation

- Non-Bayesian view of degrees of belief
- Degrees of belief should
  - reflect the amount of available supporting evidence
    $\rightarrow$ *degrees of support*
  - change non-monotonically when new evidence arrives
  - be consistent with given logical and probabilistic constraints
- Consequently, the complete absence of evidence should always imply zero degrees of belief/support
- Thus, degrees of belief/support of complementary hypothesis will not always add up to one

  $\rightarrow$ *sub-additivity*
- How can we get sub-additivity in a probabilistic calculus without violating Kolmogorov's axioms?

Example

- Alice's barbecue party:

  > *"Alice flips a fair coin and promises to organize a barbecue tomorrow night if the coin lands on head. Alice is well known to always keep her promises, but she does not say anything about what she is doing if the coin lands on tail, i.e. she may or may not organize the barbecue. That's all you know about Alice and her barbecue."*

- Degree of support for the barbecue to take place? Degree of support for the barbecue to be canceled?
  - $dsp(B) = 0.5$
  - $dsp(\neg B) = 0$
- Degree of *possibility* for the barbecue to take place?
  - $dps(B) = 1 - dsp(\neg B) = 1$

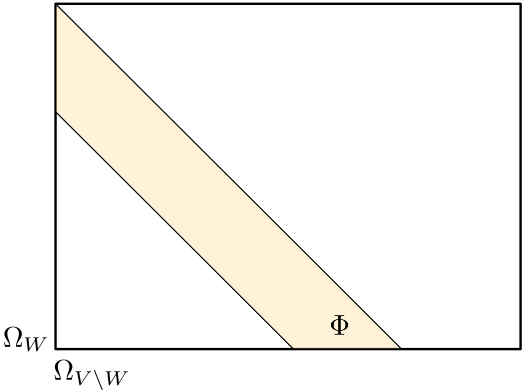
Formal Framework I

- We presuppose a logical language $\mathcal{L}_V$ defined over a set of variables V (usually discrete)

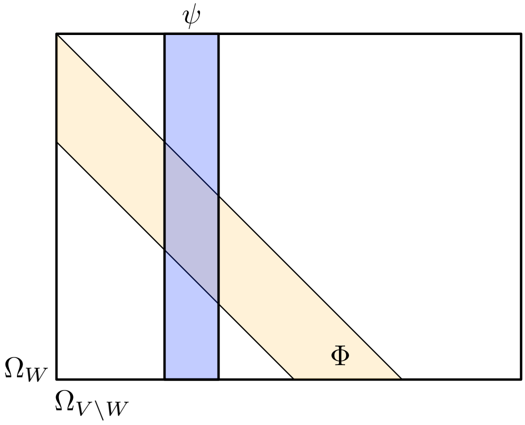
$$\begin{aligned}
\Phi \subset \mathcal{L}_V &\Rightarrow \text{background knowledge (evidence)}\\
W \subseteq V &\Rightarrow \text{“probabilistic variables”}\\
\Omega_W &\Rightarrow \text{possible states w.r.t. } W \text{ (called \emph{scenarios})}\\
P &\Rightarrow \text{probability measure over the } \sigma\text{-algebra } 2^{\Omega_W}\\
\psi \in \mathcal{L}_V &\Rightarrow \text{hypothesis (event)}\\
\text{Args}(\psi) &\Rightarrow \text{set of scenarios } \mathbf{s} \in \Omega_W \text{ such that } \Phi_{\mathbf{s}} \models \psi
\end{aligned}$$

- The elements of $\text{Args}(\psi)$ and $\text{Args}(\neg\psi)$ are called *arguments* and *counter-arguments*, respectively
- The elements of $\text{Args}(\bot)$ are called *conflicts*

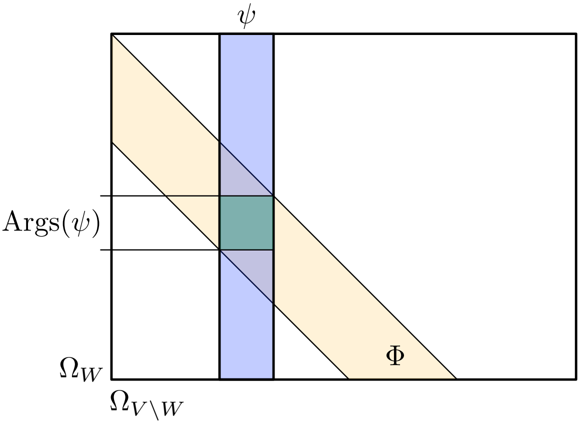
Formal Framework II



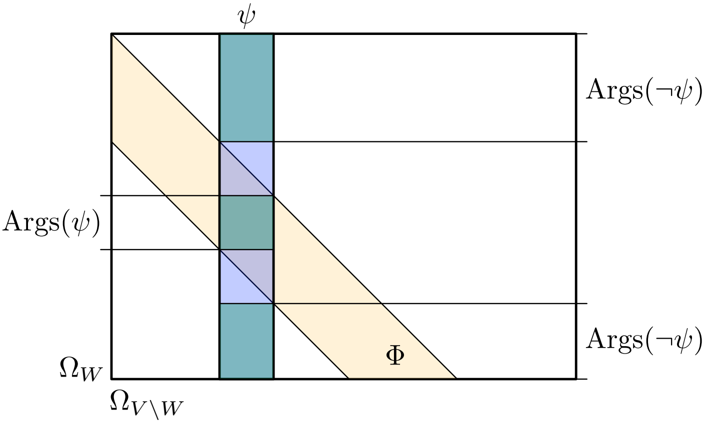
Formal Framework II



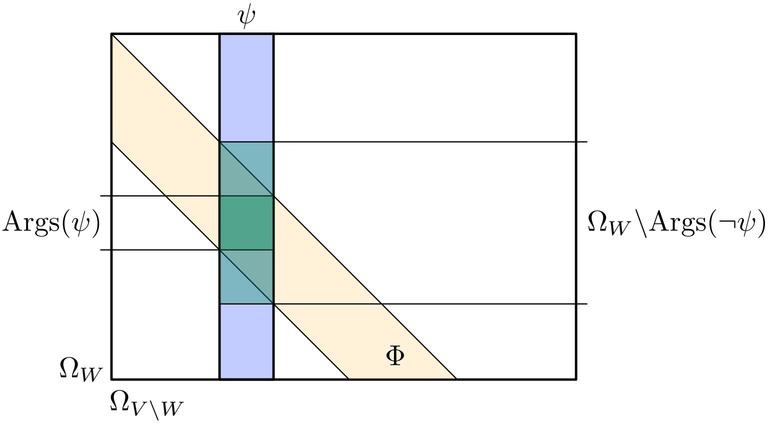
Formal Framework II



Formal Framework II



Formal Framework II



Formal Framework III

Definition: Degrees of support and possibility

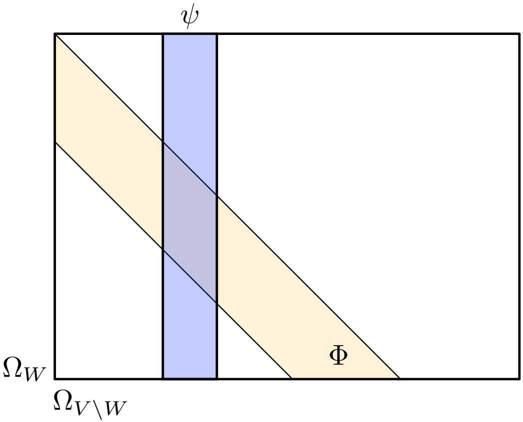
$$dsp(\psi) = P(\text{Args}(\psi) \,|\, \Omega_W \backslash \text{Args}(\bot)) = \frac{P(\text{Args}(\psi)) - P(\text{Args}(\bot))}{1 - P(\text{Args}(\bot))}$$

$$dps(\psi) = 1 - dsp(\neg\psi), \text{ i.e. } dsp(\psi) \leq dps(\psi)$$

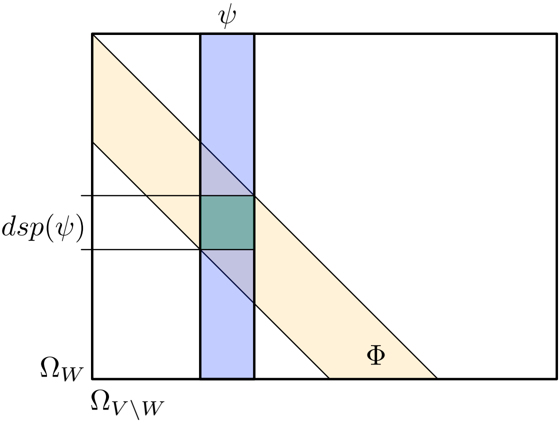
Thus, degrees of support are *ordinary* probabilities (in the sense of Kolmogorov) of *unordinary* events Args(ψ), within which ψ is a logical consequence of Φ

- sub-additive (w.r.t. ψ, but not w.r.t Args(ψ))
- non-monotone
- consistent with logical inference for $W = \emptyset$
- consistent with probabilistic inference for $W = V$

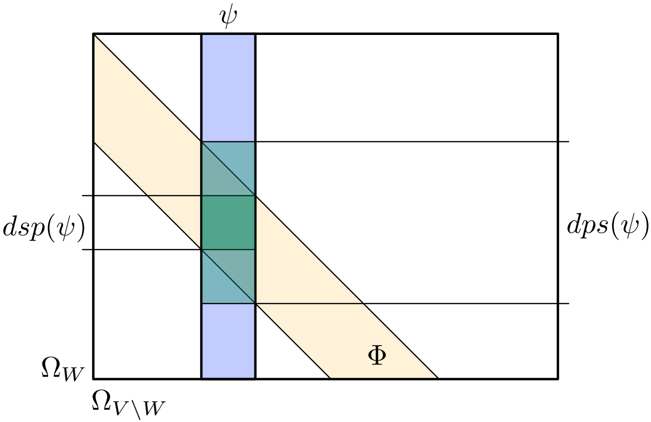
Formal Framework IV



Formal Framework IV



Formal Framework IV



Outline

Part II: Probabilistic Argumentation

Different Semantics I

Probabilistic argumentation allows many different interpretations for a given question of the form $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^{Y}$?

- Semantics 1:
  - let $\mathbb{S}$ denote all possible degree of support functions
  - let every set X_i define a constraint $dsp(\varphi_i) \in X_i$
  - consider the combined constraint $\mathbb{S}_*$
  - take $Y = \{dsp(\psi) : dsp \in \mathbb{S}_*\}$ or $Y = [\underline{dsp}(\psi), \overline{dps}(\psi)]$
- Semantics 2:
  - consider a subset of variables $W \subseteq Vars(\{\varphi_1, \ldots, \varphi_n\})$
  - let $\mathbb{P}$ denote all possible probability measures w.r.t. W
  - let every set X_i define a constraint $P(\varphi_i^{\downarrow W}) \in X_i$
  - consider the combined constraint $\mathbb{P}_*$
  - take $Y = \{dsp(\psi) : P \in \mathbb{P}_*\}$ or $Y = [\underline{dsp}(\psi), \overline{dps}(\psi)]$

Different Semantics II

- Semantics 3:
  - let $X_i = \{\ell_i, u_i\}$ be an interval
  - let $\mathbb{S}$ denote all possible degree of support functions
  - let every ℓ_i define a constraint $dsp(\varphi_i) = \ell_i$
  - let u_i define a constraint $dps(\varphi_i) = 1 - dsp(\neg\varphi_i) = u_i$
  - consider the combined constraint $\mathbb{S}_*$
  - take $Y = \{dsp(\psi) : dsp \in \mathbb{S}_*\}$ or $Y = [\underline{dsp}(\psi), \overline{dps}(\psi)]$
- Semantics 4:
  - let $X_i = \{x_i\}$ be a sharp value
  - let x_i represent the *evidential uncertainty* of φ_i, e.g. look at φ_i as a statement of an unreliable source S_i with $P(rel_i) = x_i$
  - let $W = \{rel_1, \ldots, rel_n\}$, $\Phi = \{rel_1 \rightarrow \varphi_1, \ldots, rel_n \rightarrow \varphi_n\}$
  - assuming mutually independent sources defines a fully specified probability measure over W
  - take $Y = [dsp(\psi), dps(\psi)]$

Different Semantics III

- Semantics 5:
  - let $X_i = \{\ell_i, u_i\}$ be an interval
  - look at φ_i as the statement of a possibly unreliable source S_i
  - interpret "unreliable" as "incompetent or dishonest", i.e. "reliable" is interpreted as "competent and honest"
  - assume independence between $comp_i$ and hon_i
  - define $P(comp_i) = 1 - (u_i - \ell_i)$ and $P(hon) = \frac{1-\ell_i}{1-(u_i-\ell_i)}$
  - let $W = \{comp_1, hon_1 \ldots, comp_n, hon_n\}$
  - assuming mutually independent sources defines a fully specified probability measure over W
  - $\Phi = \{comp_1 \rightarrow (hon_1 \leftrightarrow \varphi_1), \ldots, comp_n \rightarrow (hon_n \leftrightarrow \varphi_n)\}$
  - take $Y = [dsp(\psi), dps(\psi)]$
- and many more ...

Outline

Part III: Probabilistic Networks

5. Bayesian and Credal Networks

6. Computational Methods for Credal Networks

Outline

Part III: Probabilistic Networks

5 Bayesian and Credal Networks

6 Computational Methods for Credal Networks

The Progicnet Strategy

- All approaches under the umbrella of the Progic framework require some sorts of probability sets
- Generally, computations with such sets of probabilities is very complicated and complex
- Probabilistic networks help to reduce the computational complexity of probabilistic inference
  - Bayesian networks (for single probability functions)
  - Credal networks (for sets of probability functions)
- The Progicnet strategy consists in using credal networks as a common computational machinery
  - promising preliminary results for some of the possible semantics
  - work in progress ...

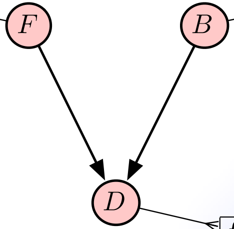
Bayesian Networks

- Bayesian and credal networks are similar
  - the network forms a DAG
  - a variable $X \in \mathbf{X}$ it associated to each network node
  - arrows represent conditional independencies among variables
  - observed evidence $\mathbf{E} = \mathbf{e}$, for evidence variables $\mathbf{E} \subseteq \mathbf{X}$
  - hypothesis $H = h$, for query variable $H \in \mathbf{X}$
- Inference in Bayesian networks
  - conditional probabilities: $P(X|parents(X))$
  - joint probability functions: $P(\mathbf{X}) = \prod_{X \in \mathbf{X}} P(X|parents(X))$
  - posterior probabilities: $P(h|\mathbf{e}) = \dfrac{P(h, \mathbf{e})}{P(\mathbf{e})}$

Example

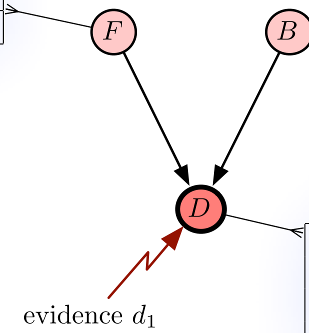
$P(F)$	
f_1	f_2
0.15	0.85

$P(B)$	
b_1	b_2
0.01	0.99



$P(D \mid F, B)$	d_1	d_2
f_1, b_1	0.99	0.01
f_1, b_2	0.97	0.03
f_2, b_1	0.9	0.1
f_2, b_2	0.3	0.7

Example

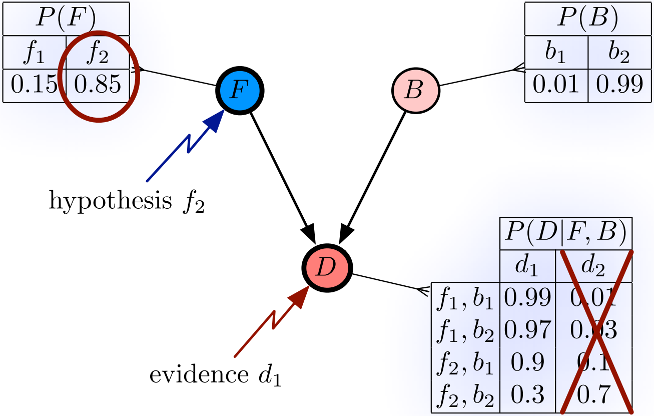


$P(F)$	
f_1	f_2
0.15	0.85

$P(B)$	
b_1	b_2
0.01	0.99

$P(D\|F,B)$	d_1	~~d_2~~
f_1, b_1	0.99	~~0.01~~
f_1, b_2	0.97	~~0.03~~
f_2, b_1	0.9	~~0.1~~
f_2, b_2	0.3	~~0.7~~

Example

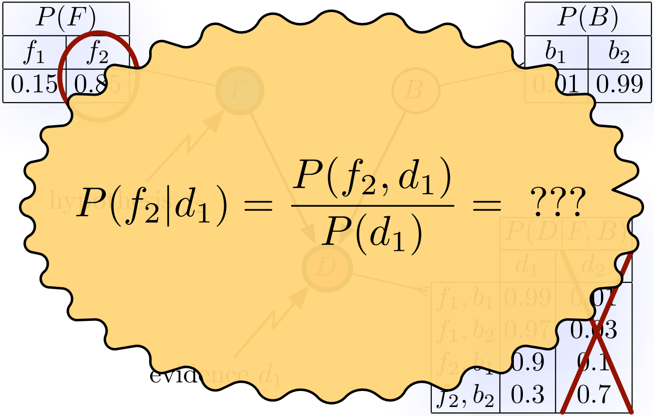


$P(F)$	
f_1	f_2
0.15	0.85

$P(B)$	
b_1	b_2
0.01	0.99

	$P(D\|F,B)$	
	d_1	d_2
f_1, b_1	0.99	0.01
f_1, b_2	0.97	0.03
f_2, b_1	0.9	0.1
f_2, b_2	0.3	0.7

Example



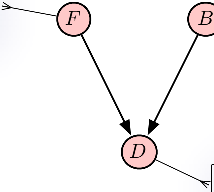
$$P(f_2|d_1) = \frac{P(f_2, d_1)}{P(d_1)} = \;???$$

Credal Networks

- A credal network relaxes the uniqueness assumptions for the given probability values
- Probability functions are replaced by credal sets
  - $K(X)$ = closed convex set of probability functions $P(X)$
  - $\text{Ext}(K(X)) = \{P_1(X), \ldots, P_m(X)\}$ = extremal points
- Inference in credal networks
  - conditional credal sets: $K(X|parents(X))$
  - largest joint credal set: $K(\mathbf{X})$
  - lower posterior probabilty: $\underline{P}(h|\mathbf{e})$
  - upper posterior probabilty: $\overline{P}(h|\mathbf{e})$
- The *extension* of a credal network determines independence assumptions that the members of the credal sets satisfy
  - natural extension: no independent assumptions
  - strong extension: independence for the extremal points

Example

$\text{Ext}(K(F))$			
$P_1(F)$		$P_2(F)$	
f_1	f_2	f_1	f_2
0.15	0.85	0.25	0.75



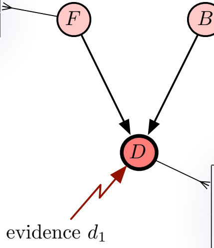
$\text{Ext}(K(B))$			
$P_1(B)$		$P_2(B)$	
b_1	b_2	b_1	b_2
0.01	0.99	0.05	0.95

	$\text{Ext}(K(D\|F,B))$			
	$P_1(D\|F,B)$		$P_2(D\|F,B)$	
	d_1	d_2	d_1	d_2
f_1, b_1	0.99	0.01	0.95	0.05
f_1, b_2	0.97	0.03	0.99	0.01
f_2, b_1	0.9	0.1	1	0
f_2, b_2	0.3	0.7	0.5	0.5

Example

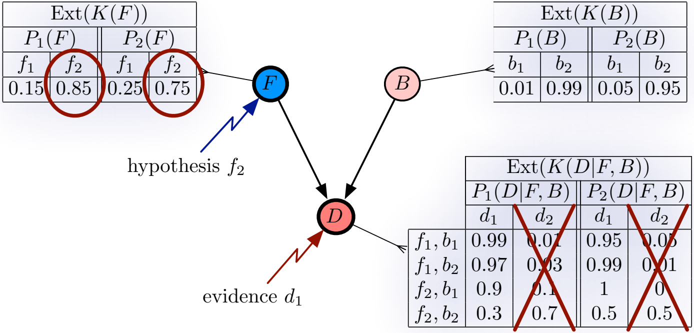
Ext($K(F)$)			
$P_1(F)$		$P_2(F)$	
f_1	f_2	f_1	f_2
0.15	0.85	0.25	0.75

Ext($K(B)$)			
$P_1(B)$		$P_2(B)$	
b_1	b_2	b_1	b_2
0.01	0.99	0.05	0.95



	Ext($K(D\|F,B)$)			
	$P_1(D\|F,B)$		$P_2(D\|F,B)$	
	d_1	~~d_2~~	d_1	~~d_2~~
f_1, b_1	0.99	~~0.01~~	0.95	~~0.05~~
f_1, b_2	0.97	~~0.03~~	0.99	~~0.01~~
f_2, b_1	0.9	~~0.1~~	1	~~0~~
f_2, b_2	0.3	~~0.7~~	0.5	~~0.5~~

Example



Ext($K(F)$)			
$P_1(F)$		$P_2(F)$	
f_1	f_2	f_1	f_2
0.15	0.85	0.25	0.75

Ext($K(B)$)			
$P_1(B)$		$P_2(B)$	
b_1	b_2	b_1	b_2
0.01	0.99	0.05	0.95

	Ext($K(D\|F,B)$)			
	$P_1(D\|F,B)$		$P_2(D\|F,B)$	
	d_1	d_2	d_1	d_2
f_1, b_1	0.99	0.01	0.95	0.05
f_1, b_2	0.97	0.03	0.99	0.01
f_2, b_1	0.9	0.1	1	0
f_2, b_2	0.3	0.7	0.5	0.5

Example

Ext($K(F)$)			
$P_1(F)$		$P_2(F)$	
f_1	f_2	f_1	f_2
0.15	0.85	0.25	0.75

Ext($K(B)$)			
$P_1(B)$		$P_2(B)$	
b_1	b_2	b_1	b_2
0.01	0.99	0.05	0.95

$$\overline{P}(f_2|d_1) = \ ???$$

$$\underline{P}(f_2|d_1) = \ ???$$

hypothesis f_2

evidence d_1

	Ext($K(D\|F,B)$)			
	$P_1(D\|F,B)$		$P_2(D\|F,B)$	
	d_1	d_2	d_1	d_2
f_1,b_1	0.99	0.01	0.95	0.05
f_1,b_2	0.97	0.03	0.99	0.01
f_2,b_1	0.9	0.1	1	0
f_2,b_2	0.3	0.7	0.5	0.5

Parametrised Credal Network

- A parametrised credal network represents a credal set in which the extremal points are interrelated
- In the Progic framework, such relations may arise when the constraints involve more than one network node
- Example
  - $a^{[0.3,1]}$, i.e. $\gamma = P(a) \in [0.3, 1]$
  - $(a \wedge b)^{\{0.2\}}$, i.e. $P(a \wedge b) = 0.2$
  - $P(b|a) = \frac{0.2}{\gamma}$
- If the functional relations between the interval bounds respect certain restrictions, parametrised credal networks offer the same computational advantages as ordinary credal sets

The Progic Strategy (revisited)

- The inference problem in the Progic framework is to find minimal sets Y such that $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^Y$
- The general idea is to use (parametrised) credal networks to make inference tractable
- Step 1:
  - Work out the specifics (e.g. independence assumptions) of a particular semantics for the Progic framework
  - Use these specifics to build up a probabilistic network
- Step 2:
  - Use the network from Step 1 to determine Y efficiently
  - This step is independent of the chosen semantics

Outline

Part III: Probabilistic Networks

Inference Methods for Credal Networks

- Exact inference in credal networks is NP^{PP}-complete (terribly complex), and NP-complete for a bounded treewidth
- Several good approximation methods exist
- To meet the the requirements of the Progic framework, such an algorithm must be able to cope with complex hypotheses
- One strategy is to transform the hypothsis ψ into a disjoint DNF $\psi_1 \vee \cdots \vee \psi_r$ for which $\psi_i \wedge \psi_j \equiv \bot$ if $i \neq j$
- This implies $P(\psi) = P(\psi_1) + \cdots + P(\psi_r)$
- Example:

$$\psi = (a \vee b) \wedge (a \vee c) \equiv a \vee (b \wedge c) \equiv a \vee (\neg a \wedge b \wedge c)$$

- Note that $\overline{P}(\psi) = \overline{P}(\psi_1 \vee \cdots \vee \psi_r) \neq \overline{P}(\psi_1) + \cdots + \overline{P}(\psi_r)$

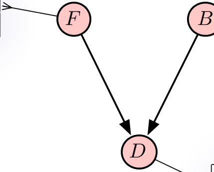
Hill-Climbing on Compiled Credal Networks

- A suitable method has been developed within Progicnet
  - R. Haenni, "*Climbing the Hills of Compiled Credal Networks*", ISIPTA, 2007
- Step 1: Logical compilation (offline)
  - represent the network structure logically as a d-DNNF
  - possibly expensive, but only required once
- Step 2: Add evidence/hypothesis
  - for some given evidence and a hypothesis, adapt the d-DNNF from Step 1 accordingly
  - cheap
- Step 3: Hill-climbing algorithm
  - use the result from Step 2 to perform the hill-climbing algorithm (steepest ascent, random restart)
  - each hill-climbing step is cheap

Example

Ext($K(F)$)			
$P_1(F)$		$P_2(F)$	
f_1	f_2	f_1	f_2
0.15	0.85	0.25	0.75

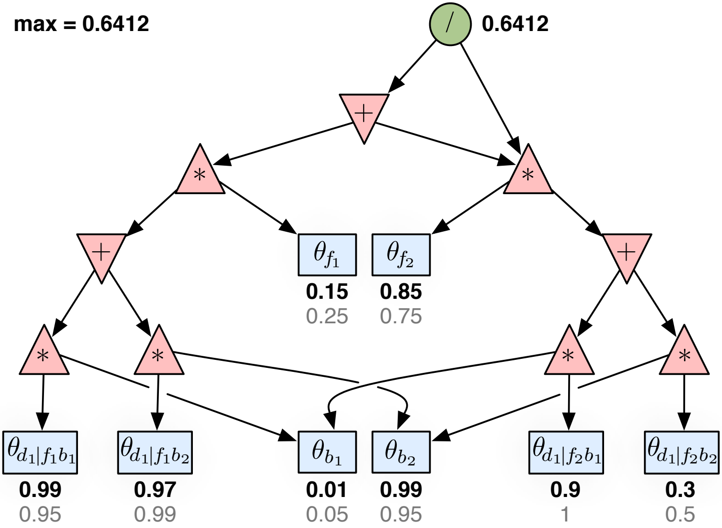
Ext($K(B)$)			
$P_1(B)$		$P_2(B)$	
b_1	b_2	b_1	b_2
0.01	0.99	0.05	0.95



	Ext($K(D\|F,B)$)			
	$P_1(D\|F,B)$		$P_2(D\|F,B)$	
	d_1	d_2	d_1	d_2
f_1, b_1	0.99	0.01	0.95	0.05
f_1, b_2	0.97	0.03	0.99	0.01
f_2, b_1	0.9	0.1	1	0
f_2, b_2	0.3	0.7	0.5	0.5

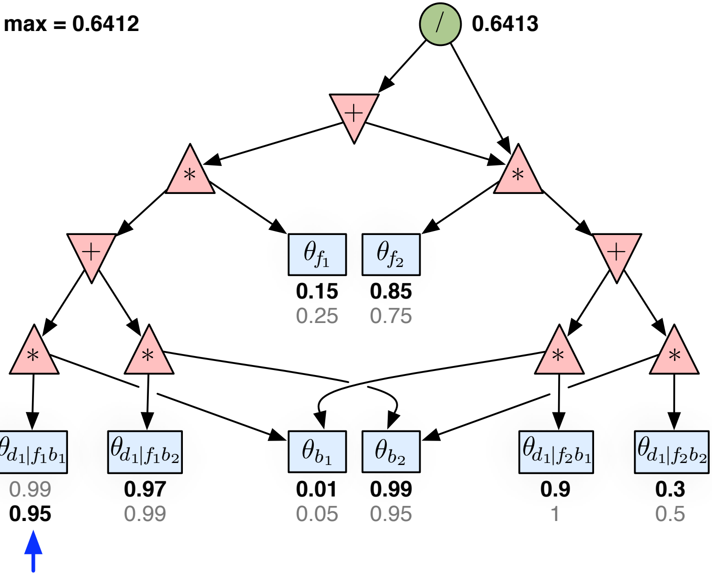
$\Rightarrow$ compute $\overline{P}(f_2|d_1)$

Example

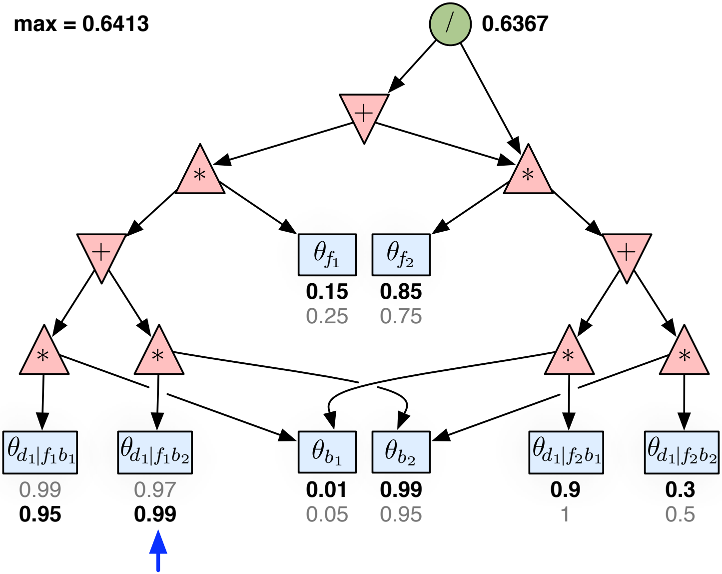


max = 0.6412
/
0.6412
+
*
*
+
θ_{f_1}
0.15
0.25
θ_{f_2}
0.85
0.75
+
*
*
*
*
$\theta_{d_1|f_1b_1}$
0.99
0.95
$\theta_{d_1|f_1b_2}$
0.97
0.99
θ_{b_1}
0.01
0.05
θ_{b_2}
0.99
0.95
$\theta_{d_1|f_2b_1}$
0.9
1
$\theta_{d_1|f_2b_2}$
0.3
0.5

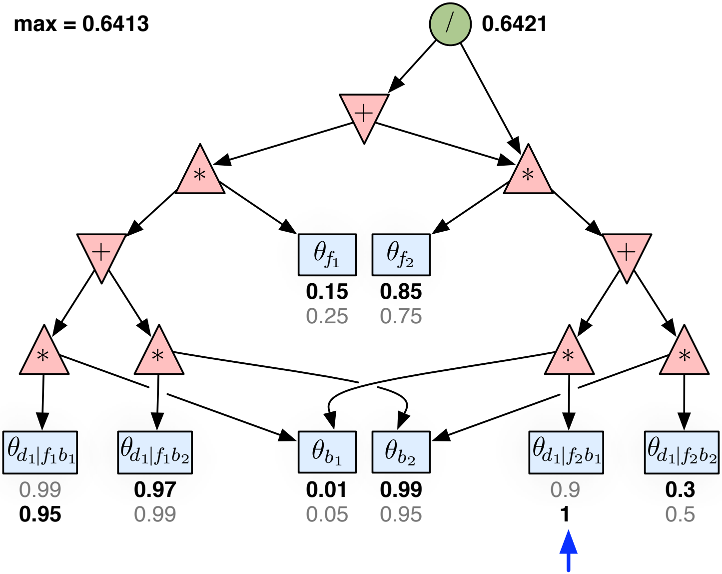
Example


max = 0.6412
/
0.6413
+
*
*
+
θ_{f_1}
0.15
0.25
θ_{f_2}
0.85
0.75
+
*
*
*
*
$\theta_{d_1|f_1b_1}$
0.99
0.95
$\theta_{d_1|f_1b_2}$
0.97
0.99
θ_{b_1}
0.01
0.05
θ_{b_2}
0.99
0.95
$\theta_{d_1|f_2b_1}$
0.9
1
$\theta_{d_1|f_2b_2}$
0.3
0.5

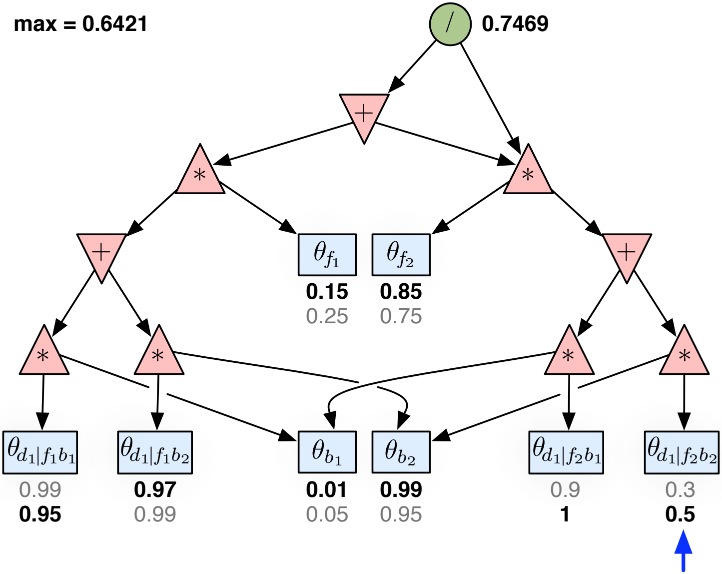
Example


max = 0.6413
/
0.6367
+
*
*
+
+
*
*
*
*
θ_{f_1}
0.15
0.25
θ_{f_2}
0.85
0.75
$\theta_{d_1|f_1b_1}$
0.99
0.95
$\theta_{d_1|f_1b_2}$
0.97
0.99
θ_{b_1}
0.01
0.05
θ_{b_2}
0.99
0.95
$\theta_{d_1|f_2b_1}$
0.9
1
$\theta_{d_1|f_2b_2}$
0.3
0.5

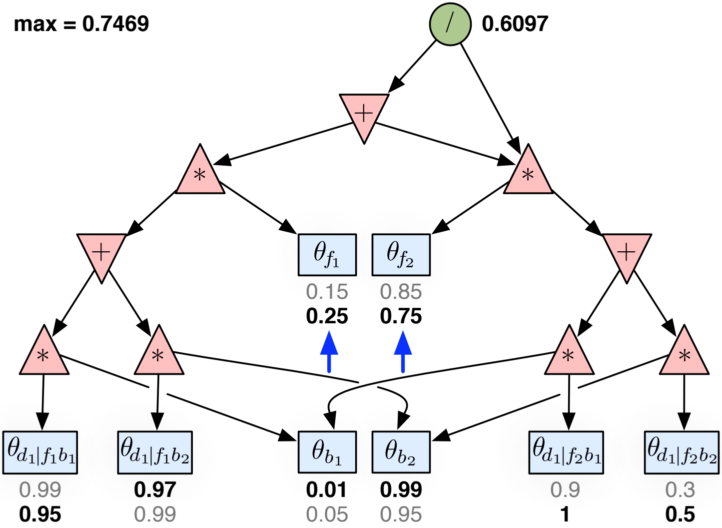
Example


max = 0.6413
/
0.6421
+
*
*
+
θ_{f_1}
θ_{f_2}
0.15
0.25
0.85
0.75
+
*
*
*
*
$\theta_{d_1|f_1b_1}$
$\theta_{d_1|f_1b_2}$
θ_{b_1}
θ_{b_2}
$\theta_{d_1|f_2b_1}$
$\theta_{d_1|f_2b_2}$
0.99
0.95
0.97
0.99
0.01
0.05
0.99
0.95
0.9
1
0.3
0.5

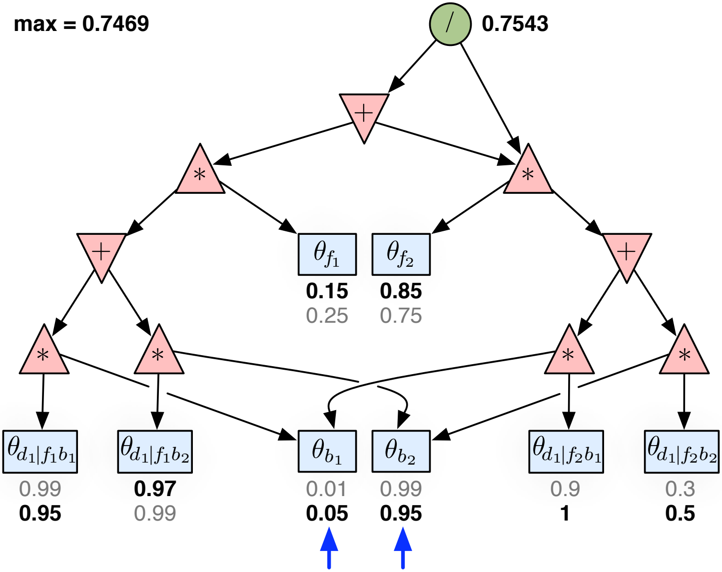
Example


max = 0.6421
/
0.7469
+
*
*
+
θ_{f_1}
0.15
0.25
θ_{f_2}
0.85
0.75
+
*
*
*
*
$\theta_{d_1|f_1b_1}$
0.99
0.95
$\theta_{d_1|f_1b_2}$
0.97
0.99
θ_{b_1}
0.01
0.05
θ_{b_2}
0.99
0.95
$\theta_{d_1|f_2b_1}$
0.9
1
$\theta_{d_1|f_2b_2}$
0.3
0.5

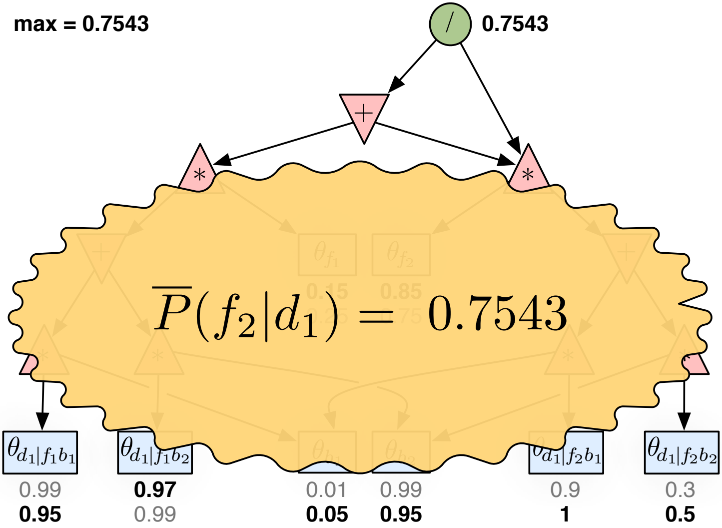
Example


max = 0.7469
/
0.6097
+
*
*
+
θ_{f_1}
θ_{f_2}
0.15
0.25
0.85
0.75
+
*
*
*
*
$\theta_{d_1|f_1b_1}$
$\theta_{d_1|f_1b_2}$
θ_{b_1}
θ_{b_2}
$\theta_{d_1|f_2b_1}$
$\theta_{d_1|f_2b_2}$
0.99
0.95
0.97
0.99
0.01
0.05
0.99
0.95
0.9
1
0.3
0.5

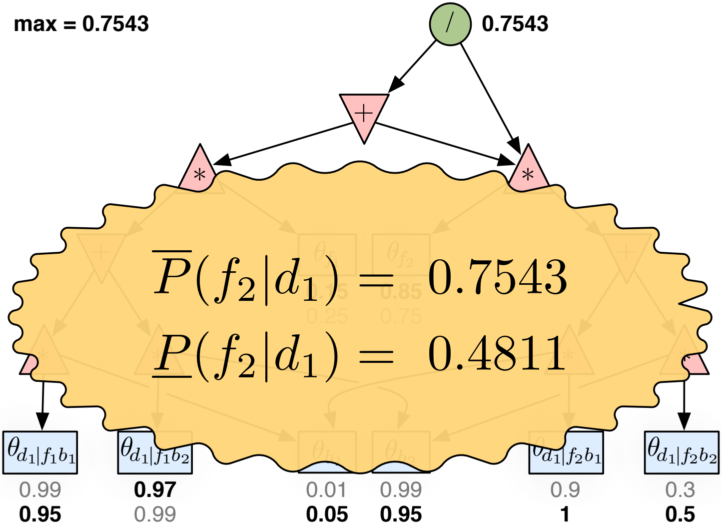
Example


max = 0.7469
/
0.7543
+
*
*
+
+
*
*
*
*
θ_{f_1}
0.15
0.25
θ_{f_2}
0.85
0.75
$\theta_{d_1|f_1b_1}$
0.99
0.95
$\theta_{d_1|f_1b_2}$
0.97
0.99
θ_{b_1}
0.01
0.05
θ_{b_2}
0.99
0.95
$\theta_{d_1|f_2b_1}$
0.9
1
$\theta_{d_1|f_2b_2}$
0.3
0.5

Example


max = 0.7543
0.7543
$\overline{P}(f_2|d_1) = 0.7543$
$\theta_{d_1|f_1b_1}$
0.99
0.95
$\theta_{d_1|f_1b_2}$
0.97
0.99
0.01
0.05
0.99
0.95
$\theta_{d_1|f_2b_1}$
0.9
1
$\theta_{d_1|f_2b_2}$
0.3
0.5

Example


max = 0.7543
/
0.7543
+
*
*
$\overline{P}(f_2|d_1) = 0.7543$
$\underline{P}(f_2|d_1) = 0.4811$
$\theta_{d_1|f_1b_1}$
0.99
0.95
$\theta_{d_1|f_1b_2}$
0.97
0.99
0.01
0.05
0.99
0.95
$\theta_{d_1|f_2b_1}$
0.9
1
$\theta_{d_1|f_2b_2}$
0.3
0.5

Conclusion

- Part I: Probabilistic Logic
  - We take the Progic framework as a common starting point
  - Find some (minimal) set Y which satisfies $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^Y$
  - This highly depends on the imposed semantics
- Part II: Probabilistic Argumentation
  - Probability space + logical evidence about some super-space
  - Leads to degrees of support and possibility
  - Allows many different semantics for the Progic framework
- Part III: Probabilistic Networks
  - Credal networks help when dealing with sets of probabilities
  - The Progicnet strategy consists in constructing such networks
  - This construction again depends on the imposed semantics, but the actual computation does not
  - Hill-climbing on the compiled network is a possible algorithm