Probabilistic Logics and Probabilistic Networks

Jan-Willem Romeijn, Philosophy, Groningen
Jon Williamson, Philosophy, Kent

ESSLLI 2008

Course Page:

- http://www.kent.ac.uk/secl/philosophy/jw/2006/progicnet/ESSLLI.htm

1 Introduction

Progicnet

Project page: http://www.kent.ac.uk/secl/philosophy/jw/2006/progicnet.htm

- Leverhulme Trust academic network 2006–8

Rolf Haenni: Computer Science and Applied Mathematics, University of Bern

Jan-Willem Romeijn: Philosophy, University of Groningen

Gregory Wheeler: Artificial Intelligence, New University of Lisbon

Jon Williamson: Philosophy, University of Kent

The Fundamental Question of Probabilistic Logic

Non-probabilistic logic.

$$\varphi_1, \varphi_2, \ldots, \varphi_n \approx \psi?$$

- Use proof methods.

Probabilistic logic. Not

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \approx \psi^{Y}?$$

but

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \approx \psi^{?}$$

- Use probabilistic inference methods.

The Progicnet Programme

Framework. A unifying framework for probabilistic logic can be constructed around entailment relationships of the form $\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!\models \psi^Y$

Standard Probabilistic Semantics: $Y = \{P(\psi) : P \text{ satisfies premisses}\}$

Probabilistic Argumentation: $Y =$ probability of worlds where entailment holds

Evidential Probability: $Y =$ risk level associated with statistical inferences

Bayesian Statistics: $Y =$ probabilities yielded by Bayes' theorem

Objective Bayesian Epistemology: $Y =$ appropriate degree of belief in ψ

Calculus. Probabilistic networks can provide a calculus for probabilistic logic—they can be used to provide answers to the fundamental question $\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!\models \psi^?$

Network Construction: Build a net to represent those P that satisfy the premisses

Inference: Calculate Y from the net

These Lectures

Reading: Haenni, Romeijn, Wheeler, Williamson: *Probabilistic Logic and Probabilistic Networks*

Info: http://www.kent.ac.uk/secl/philosophy/jw/2006/progicnet/ESSLLI.htm

Schedule.

Day	Topic	Reading
Monday	The Progicnet Programme	§§1,8
Tuesday	Standard Semantics	§§2,9
Wednesday	Evidential Probability	§§4,11
	Probabilistic Argumentation	§§3,10
	Classical Statistics	§§5,12
Thursday	Bayesian Statistics	§§6,13
Friday	Objective Bayesian Epistemology	§§7,14

Contents

2 Probabilistic Logics

The Potential of Probabilistic Logic

Useful for any application concerned with reasoning about structure in the face of uncertainty:

Bioinformatics: Probability that one molecule is present given the presence of others.

Philosophy of Science: Confirmation of logically complex theories provided by observations, experiments and other theories.

Natural Language Processing: Probability that a sentence has a particular meaning given past evidence and contextual factors.

Robotics: Finding the plan most likely to achieve the robot's goals.

Expert Systems: Most likely diagnosis given symptoms.

...

Development.

- Prehistory: de Morgan, Boole, Jevons, Keynes, Koopman, Ramsey, Jeffries.
- More recently: Carnap, Fisher, Kyburg, Jaynes, Pearl, Nilsson, Howson, Paris.

Current Situation. Probabilistic logics seem very

- disparate,
- hard to understand,
- computationally complex,
- ▸ so they are rarely applied.

Our Goals. Here we try to change things by

- presenting a unifying framework,
- separating the logic and the probability
- appealing to standard logics and standard procedures for uncertain reasoning,
- showing that complexity can often be avoided.

The Framework

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!\mid\; \psi^Y$$

- $\varphi_1, \ldots, \varphi_n, \psi$ sentences of a logical language.
- $X_1, \ldots, X_n, Y$ subsets of $[0, 1]$—normally intervals.
- Suitable for reasoning *under* uncertainty.

Entailment. For $\mid\!\approx$ to count as entailment we need notions of *models* and *satisfies* such that $\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \mid\!\approx \psi^Y$ iff all models of the LHS satisfy the RHS.

- Classical entailment: models = satisfies.
- Preferential entailment: models are preferred satisfiers.
- *Monotonic*: $\varphi_1^{X_1}, \ldots, \varphi_m^{X_m} \mid\!\approx \psi^Y \Rightarrow \varphi_1^{X_1}, \ldots, \varphi_m^{X_m}, \ldots, \varphi_n^{X_n} \mid\!\approx \psi^Y$.
- *Decomposable*: M models $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \Longleftrightarrow M$ models $\varphi_1^{X_1}$, …, M models $\varphi_n^{X_n}$.
- Decomposable $\Rightarrow$ Monotonic.

Probabilistic Logic.

- Models are probability functions.
  - Exactly which probability functions depends on the entailment relation.
- P satisfies ψ^Y iff $P(\psi) \in Y$.

Probability Functions over Languages

Propositional Languages. $\mathcal{L} = \{a_1, \ldots, a_n\}$. $S\mathcal{L}$ sentences.

- *Atomic states* $\Omega = \{\pm a_1 \wedge \cdots \wedge \pm a_n\}$.
- *Probability function* $P : S\mathcal{L} \longrightarrow \mathbb{R}$,
  1. $P(\omega) \geq 0$ for each $\omega \in \Omega$,
  2. $\sum_{\omega \in \Omega} P(\omega) = 1$,
  3. $P(\theta) = \sum_{\omega \models \theta} P(\omega)$ for $\theta \in S\mathcal{L}$.

Predicate Languages. $\mathcal{L}$ has

- constants $t_1, t_2, \ldots$ that identify each element of the domain
- finitely many predicate symbols
- atomic propositions $a_1, a_2, \ldots$
- $\mathcal{L}_n$ is the *finite* language on $t_1, \ldots, t_n$
- Atomic n-states Ω_n are atomic states of $\mathcal{L}_n$
- *Probability function* $P : S\mathcal{L} \longrightarrow \mathbb{R}$,
  1. $P(\omega_n) \geq 0$ for each ω_n,
  2. for each n, $\sum_{\omega_n \in \Omega_n} P(\omega_n) = 1$,
  3. for quantifier-free θ, $P(\theta) = \sum_{\omega_n \models \theta} P(\omega_n)$,
  4. $P(\forall x \theta(x)) = \lim_{m \to \infty} P(\bigwedge_{i=1}^m \theta(t_i))$, $P(\exists x \theta(x)) = \lim_{m \to \infty} P(\bigvee_{i=1}^m \theta(t_i))$.

3 Probabilistic Networks

Fundamental Question of Probabilistic Logic

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \mathrel{|\!\!\approx} \psi^{?}$$

- We need to find a *minimal* set Y to attach to ψ.
- Probabilistic Logic: models are probability functions.
- ▸ Use probabilistic networks to represent the models.
- ▸ Use probabilistic networks to calculate $P(\psi)$ for model P.

Bayesian Nets

A Bayesian net represents a probability function P over variables $A_1, \ldots, A_n$:

Directed Acyclic Graph: on $A_1, \ldots, A_n$.

Markov Condition: $A_i \perp\!\!\!\perp ND_i \mid Par_i$.

$$A_3 \perp\!\!\!\perp A_1 \mid A_2$$

Conditional Probability Distributions: $P(A_i|Par_i)$

$P(a_1) = 0.7$	$P(a_2\|a_1) = 0.2$	$P(a_3\|a_2) = 0.9$
	$P(a_2\|\bar{a}_1) = 0.1$	$P(a_3\|\bar{a}_2) = 0.4$

Then

$$P(A_1 \cdots A_n) = \prod_{i=1}^{n} P(A_i \mid Par_i)$$

Credal Nets

A credal net represents a *set* of probability functions over variables $A_1, \ldots, A_n$:

Directed Acyclic Graph: on $A_1, \ldots, A_n$.

Markov Condition: $A_i \perp\!\!\!\perp ND_i \mid Par_i$.

$$A_3 \perp\!\!\!\perp A_1 \mid A_2$$

Conditional Probability Distributions: Constraints $P(a_i|par_i) \in [l, u]$

$P(a_1) \in [0.7, 0.8]$	$P(a_2\|a_1) = 0.2$ $P(a_2\|\bar{a}_1) \in [0.1, 1]$	$P(a_3\|a_2) \in [0.9, 1]$ $P(a_3\|\bar{a}_2) \in [0.4, 0.45]$

Extensions.

Complete: $\{P : P$ represented by a Bayesian net compatible with the credal net$\}$

Strong: $[\{P : P$ represented by a Bayesian net compatible with the credal net$\}]$

Natural: $\{P : P$ satisfies constraints on conditional distributions in the net$\}$

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \dots, \varphi_n^{X_n} \approx\!\!\!\models \psi^?$$

Network Construction

Suppose

Convexity: the X_i are intervals,

Probabilistic Logic: according to the semantics, models are probability functions.

- One can determine a credal net representing $\{P : P \text{ is a model of } \varphi_1^{X_1}, \varphi_2^{X_2}, \dots, \varphi_n^{X_n}\}$.
  - The credal net depends on the chosen semantics.

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!| \;\psi^?$$

Inference

Determine the interval $Y = \{P(\psi) : P$ represented by the credal net$\}$.

Decomposition: write ψ in ddnf so $P(\psi) = \sum_{i=1}^{k} P(\psi_i)$

- where each ψ_i is a conjunction of literals

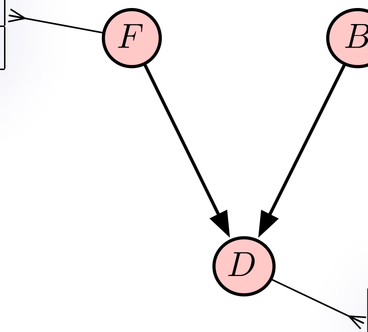
- Use the credal net to calculate $P(\psi_i)$.

Credal nets not as computationally tractable as Bayesian nets, so

Compilation Methodology: expensive offline compiling phase + cheap inference phase

- compile to a d-DNNF (deterministic Decomposable Negation Normal Form) net

Approximation Methods: hill-climbing numerical methods to approximate Y.

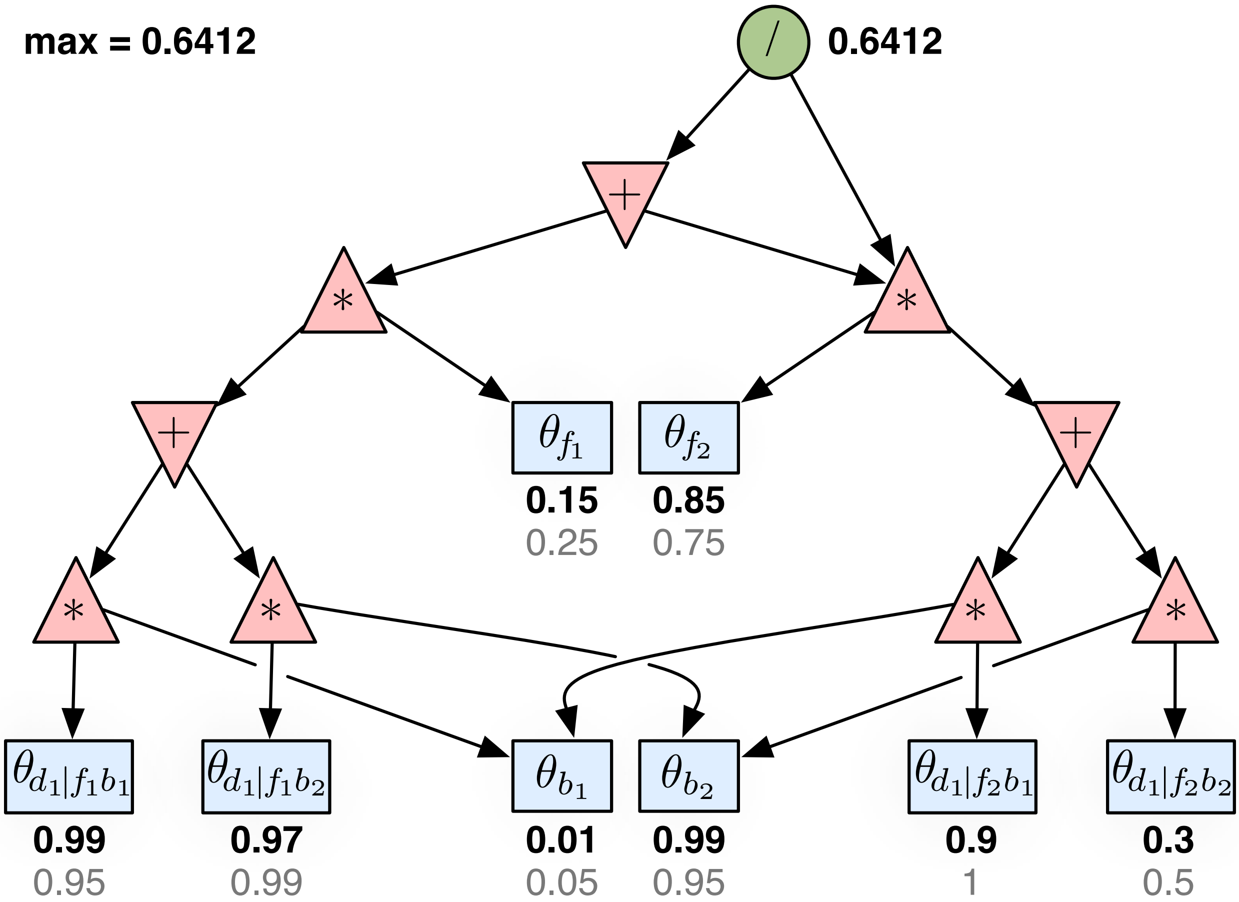


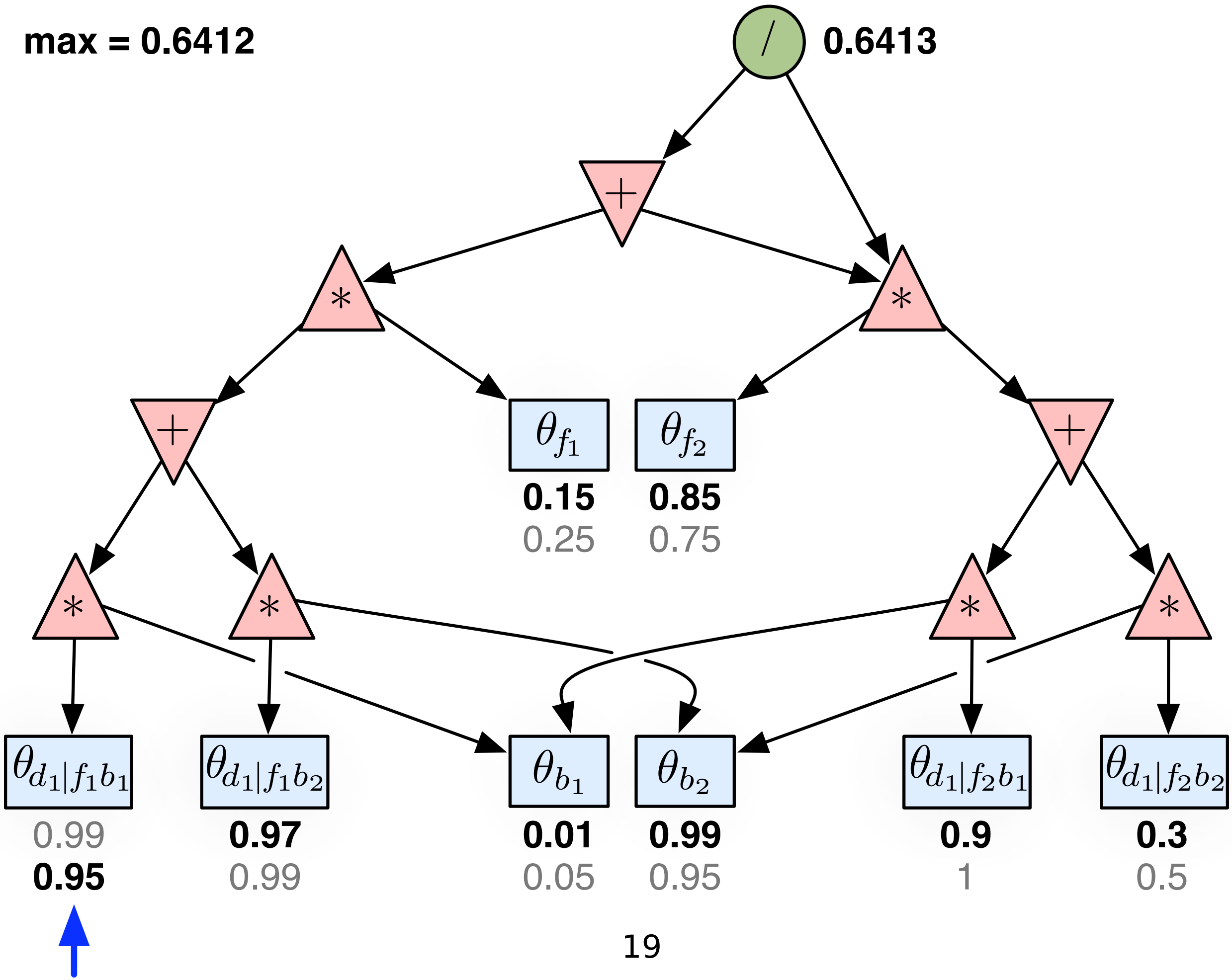
Ext($K(F)$)			
$P_1(F)$		$P_2(F)$	
f_1	f_2	f_1	f_2
0.15	0.85	0.25	0.75

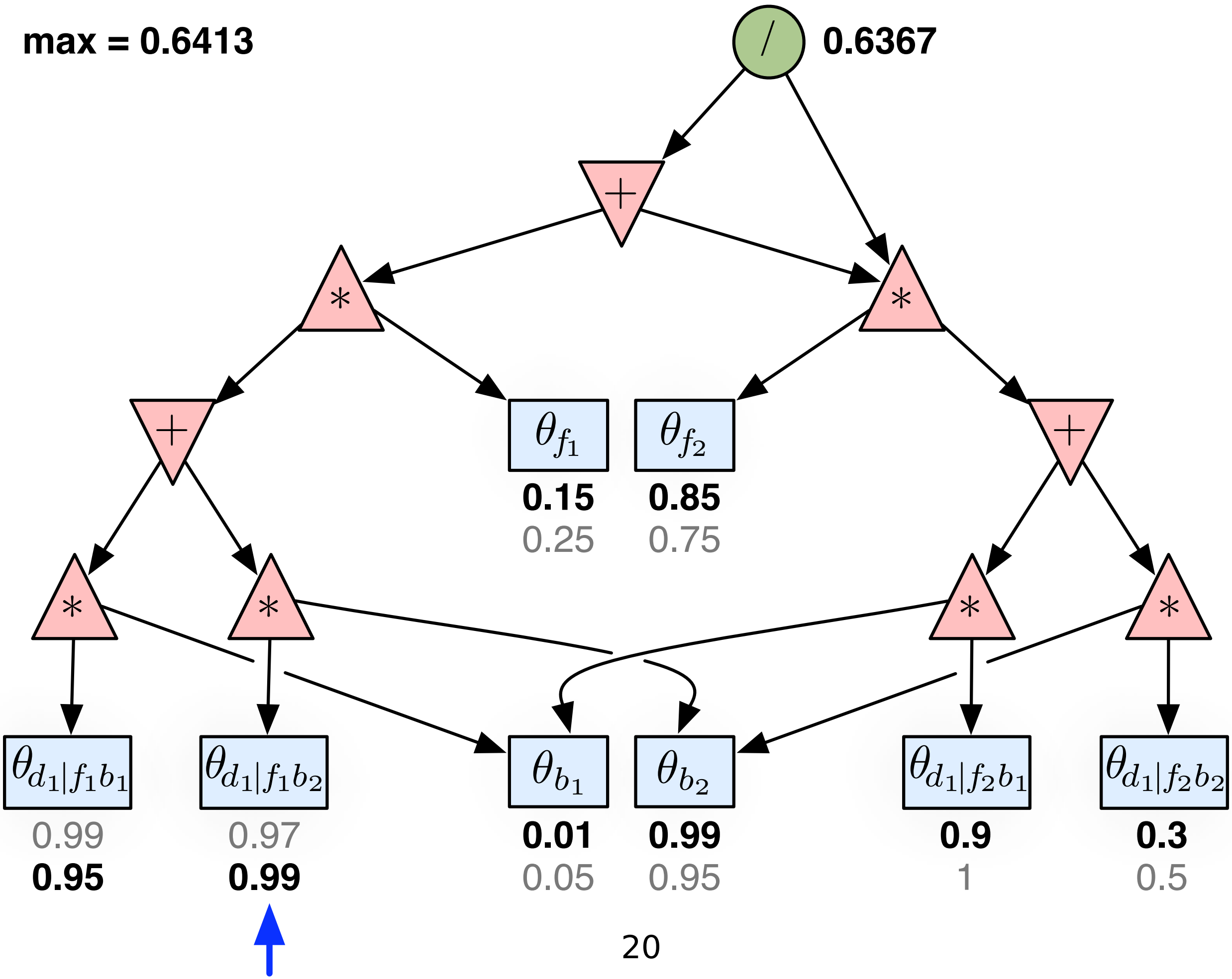
Ext($K(B)$)			
$P_1(B)$		$P_2(B)$	
b_1	b_2	b_1	b_2
0.01	0.99	0.05	0.95

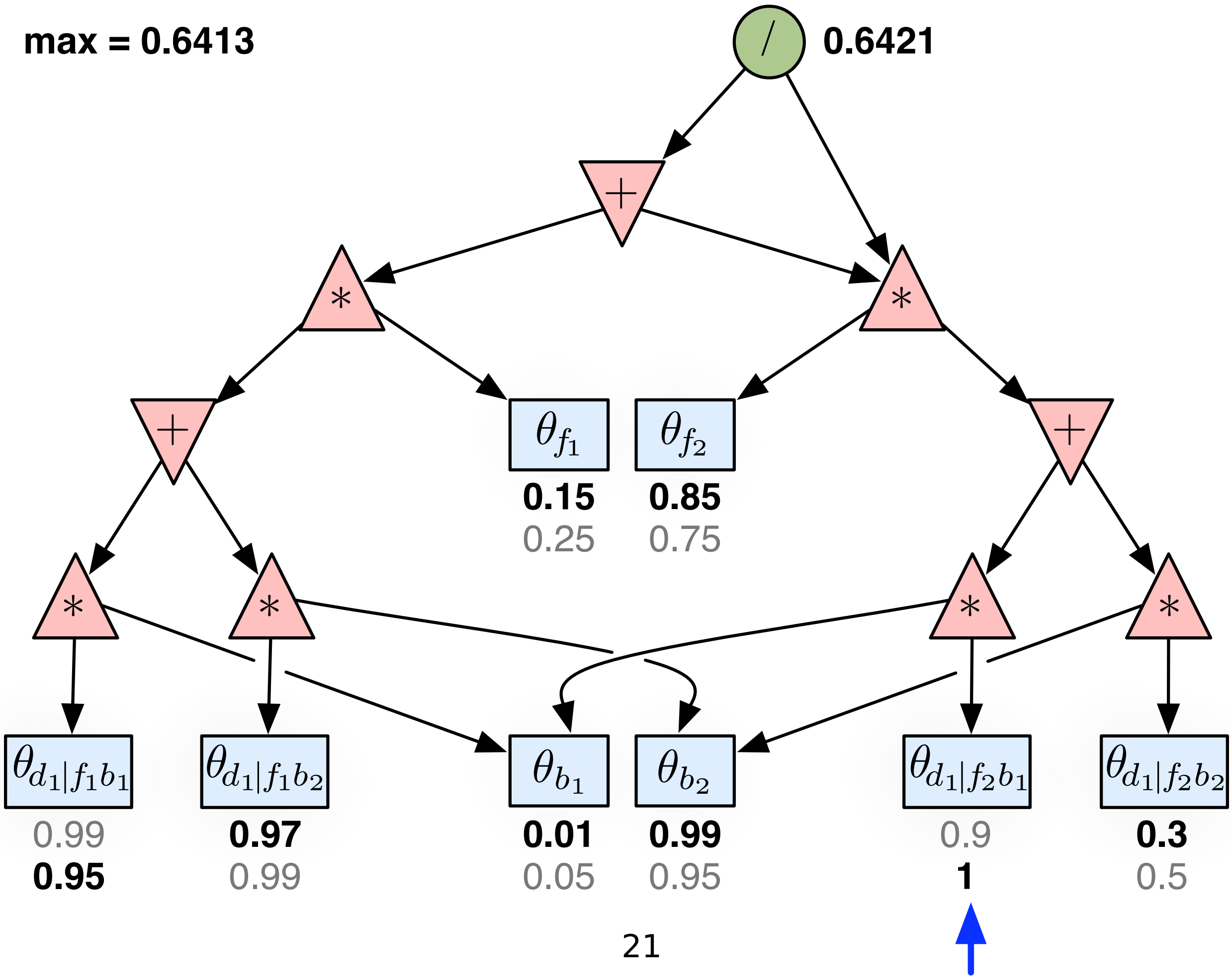
	Ext($K(D\|F,B)$)			
	$P_1(D\|F,B)$		$P_2(D\|F,B)$	
	d_1	d_2	d_1	d_2
f_1, b_1	0.99	0.01	0.95	0.05
f_1, b_2	0.97	0.03	0.99	0.01
f_2, b_1	0.9	0.1	1	0
f_2, b_2	0.3	0.7	0.5	0.5

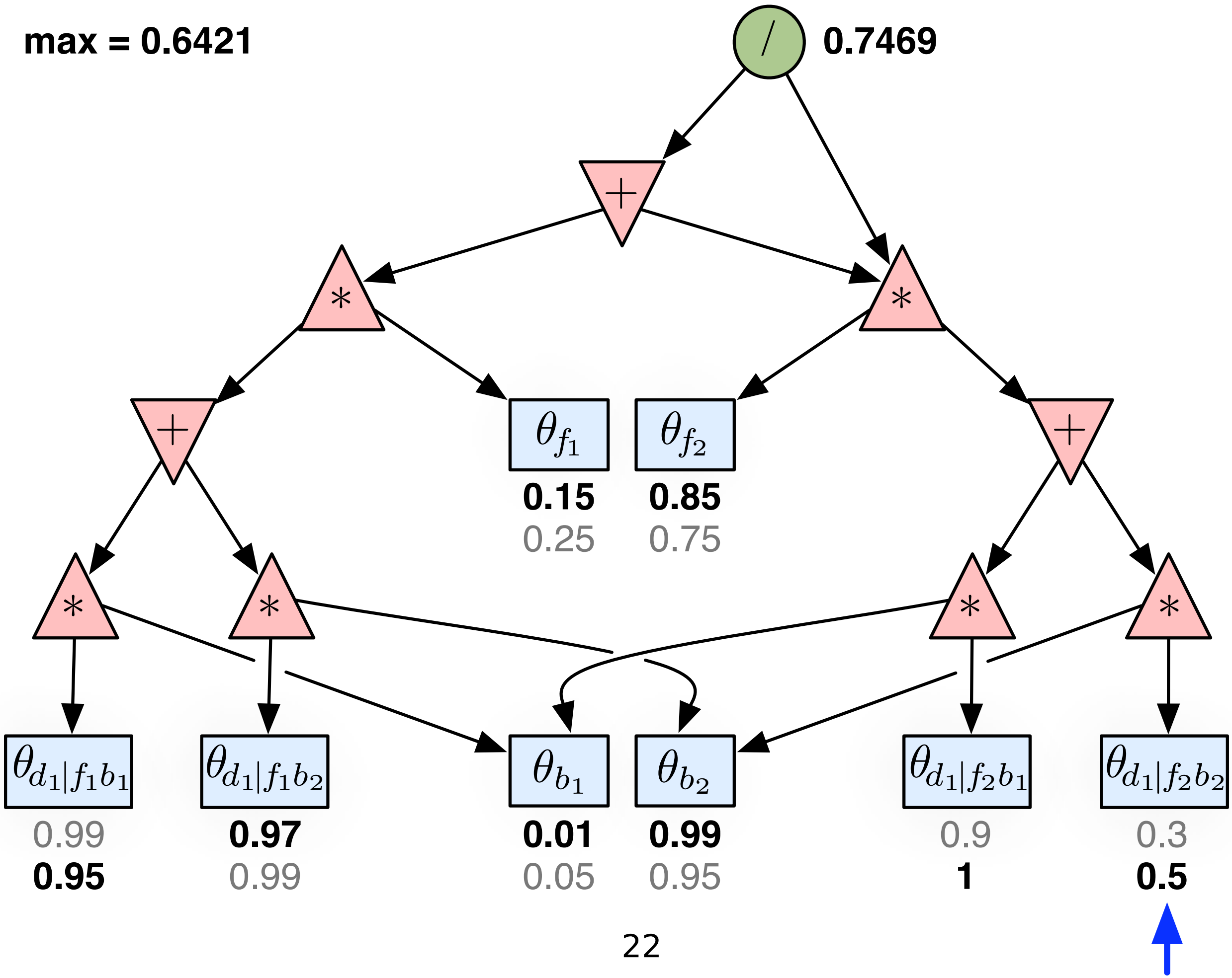
$\Rightarrow$ compute $\overline{P}(f_2|d_1)$

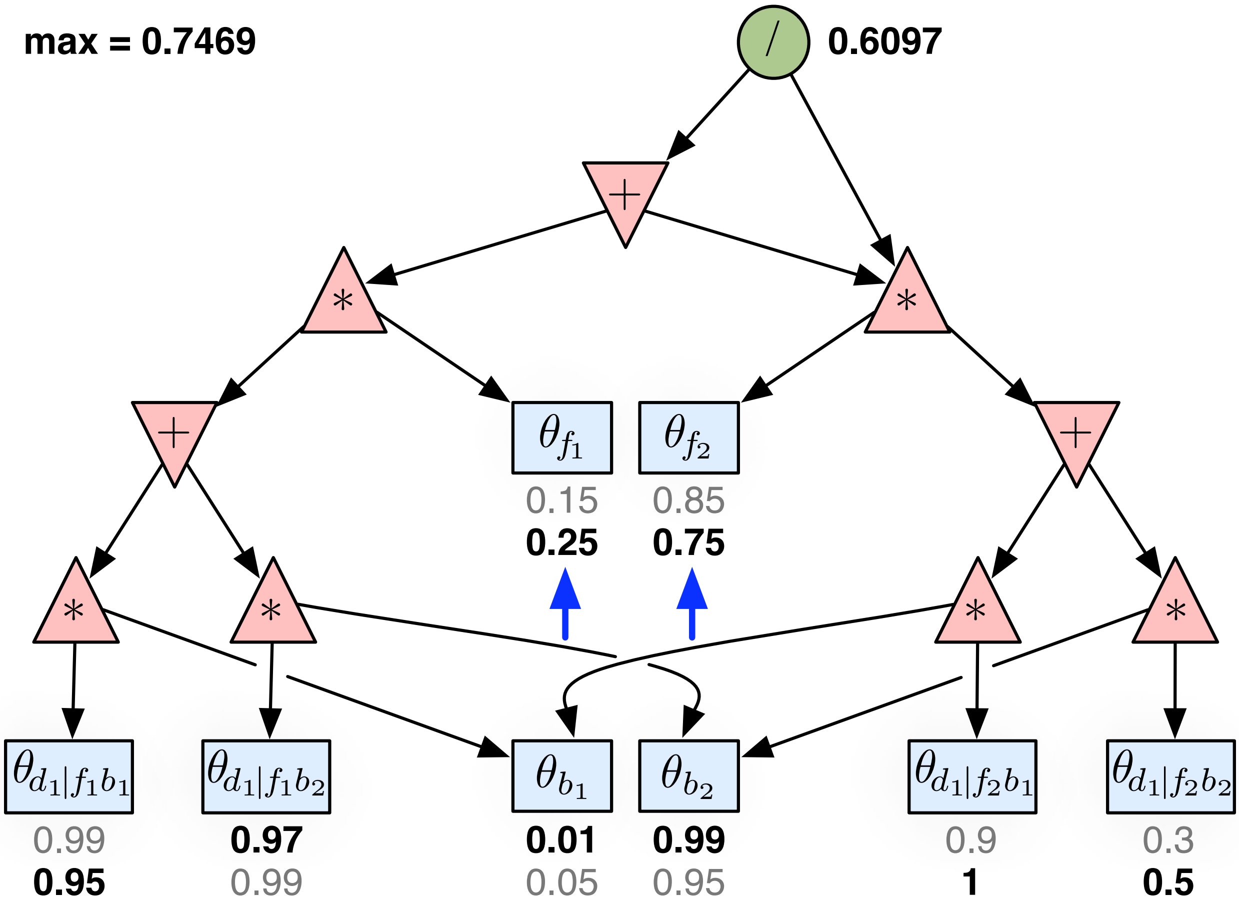

max = 0.6412
/
0.6412
+
*
*
+
θ_{f_1}
0.15
0.25
θ_{f_2}
0.85
0.75
+
*
*
*
*
$\theta_{d_1|f_1 b_1}$
0.99
0.95
$\theta_{d_1|f_1 b_2}$
0.97
0.99
θ_{b_1}
0.01
0.05
θ_{b_2}
0.99
0.95
$\theta_{d_1|f_2 b_1}$
0.9
1
$\theta_{d_1|f_2 b_2}$
0.3
0.5


max = 0.6412
/
0.6413
+
*
*
+
θ_{f1}
0.15
0.25
θ_{f2}
0.85
0.75
+
*
*
*
*
θ_{d1|f1b1}
0.99
0.95
θ_{d1|f1b2}
0.97
0.99
θ_{b1}
0.01
0.05
θ_{b2}
0.99
0.95
θ_{d1|f2b1}
0.9
1
θ_{d1|f2b2}
0.3
0.5

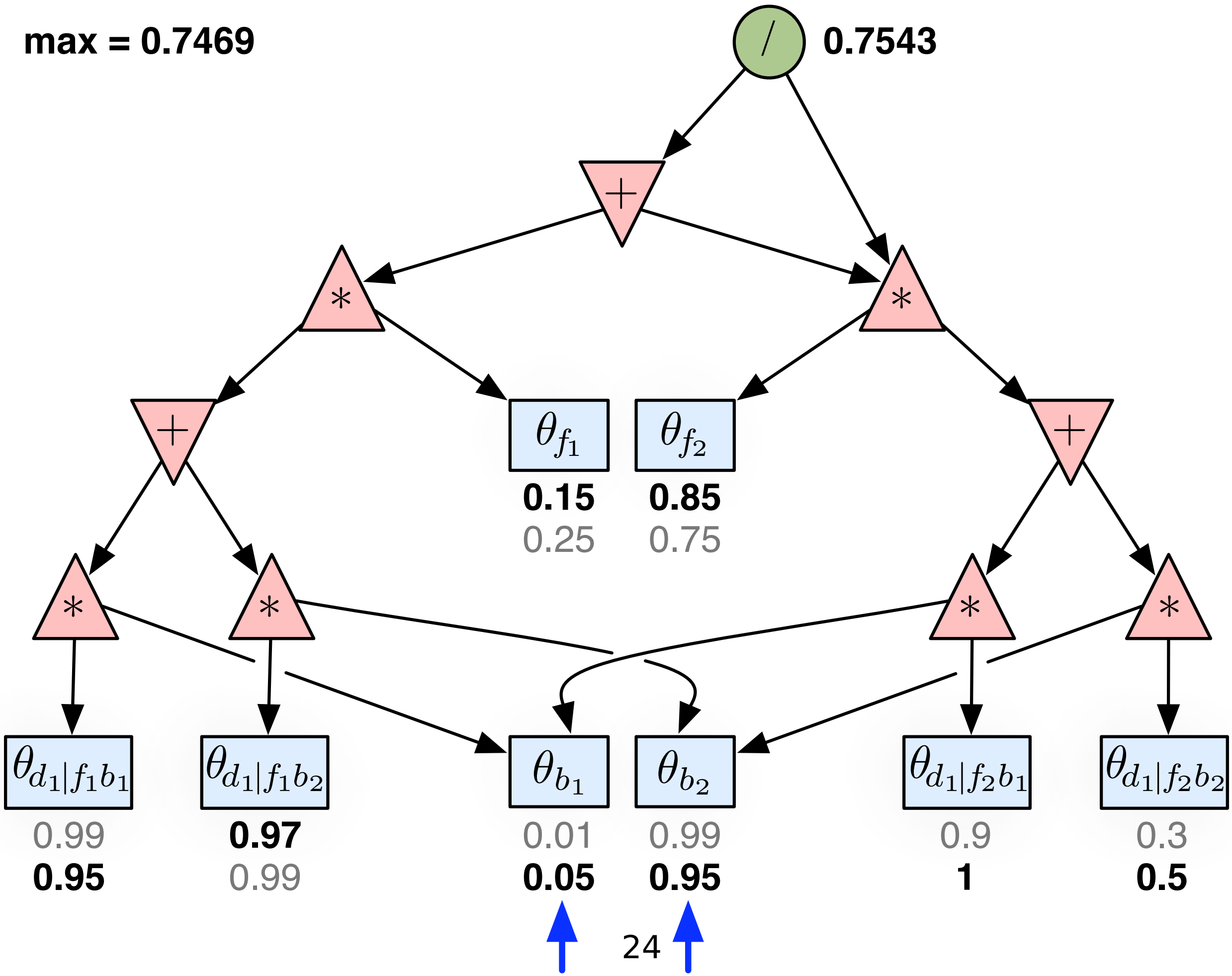

max = 0.6413
/
0.6367
+
*
*
+
θ_{f_1}
0.15
0.25
θ_{f_2}
0.85
0.75
+
*
*
*
*
$\theta_{d_1|f_1b_1}$
0.99
0.95
$\theta_{d_1|f_1b_2}$
0.97
0.99
θ_{b_1}
0.01
0.05
θ_{b_2}
0.99
0.95
$\theta_{d_1|f_2b_1}$
0.9
1
$\theta_{d_1|f_2b_2}$
0.3
0.5


max = 0.6413
/
0.6421
+
*
*
+
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

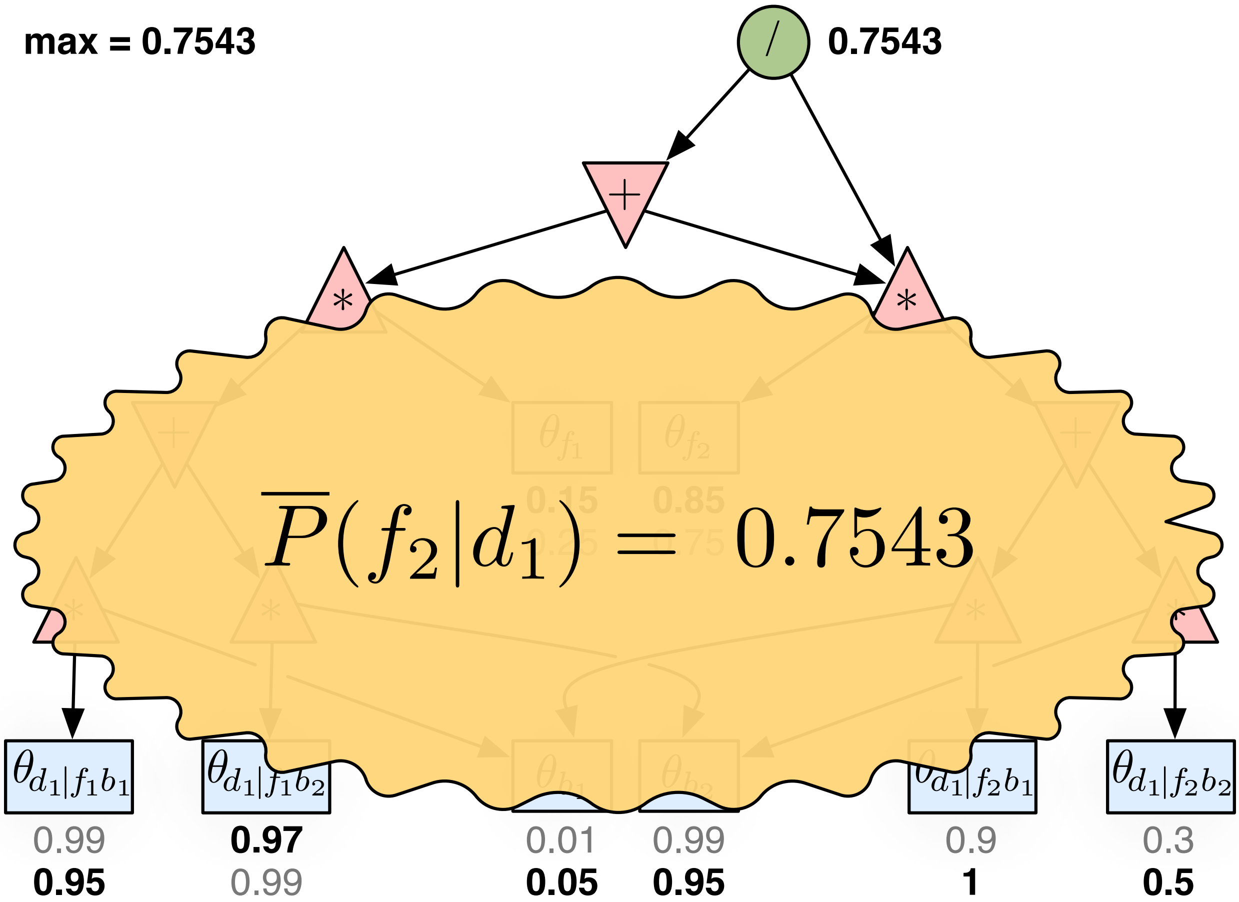

max = 0.6421
/
0.7469
+
*
*
+
θ_{f_1}
θ_{f_2}
0.15
0.25
0.85
0.75
+
*
*
*
*
$\theta_{d_1|f_1 b_1}$
0.99
0.95
$\theta_{d_1|f_1 b_2}$
0.97
0.99
θ_{b_1}
0.01
0.05
θ_{b_2}
0.99
0.95
$\theta_{d_1|f_2 b_1}$
0.9
1
$\theta_{d_1|f_2 b_2}$
0.3
0.5



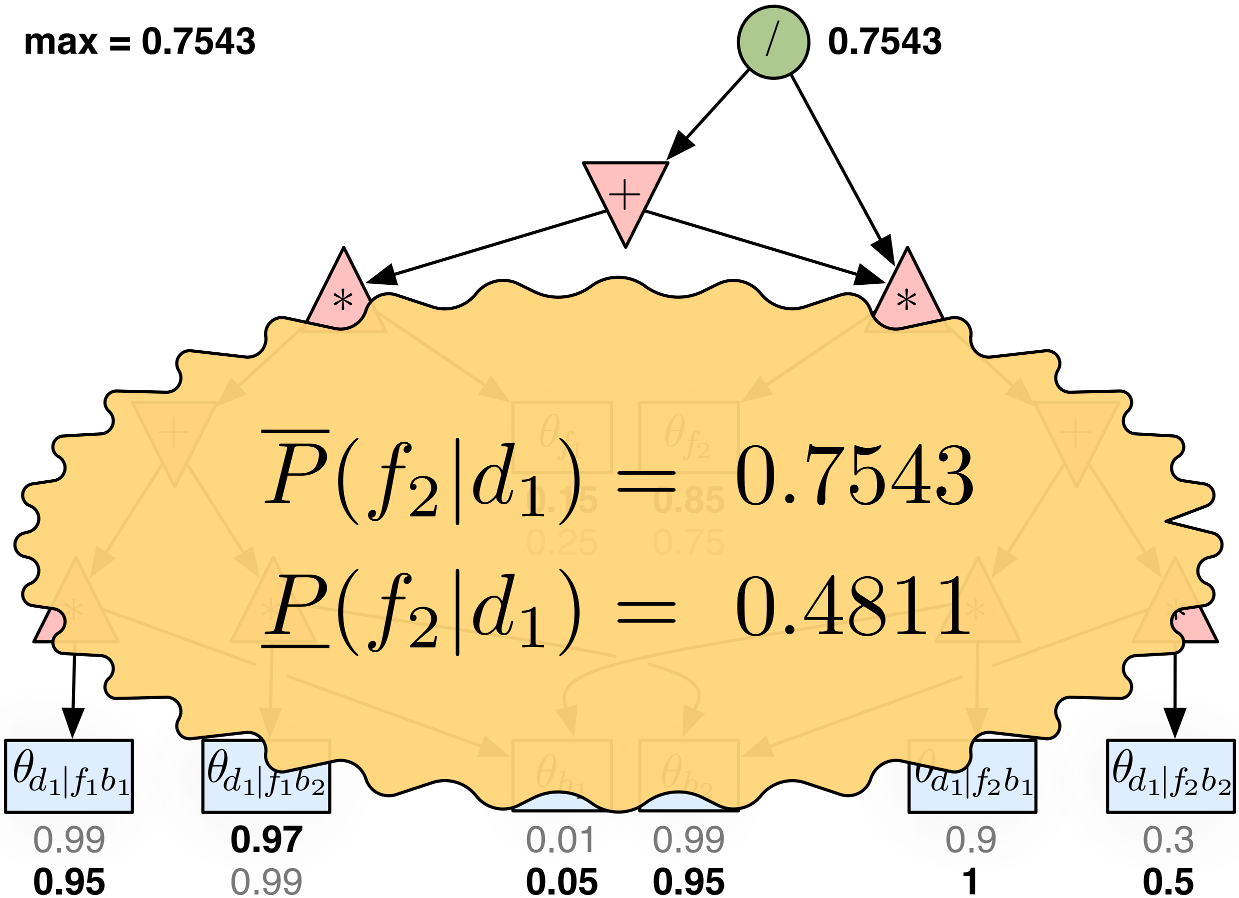
max = 0.7469
/
0.6097
+
*
*
+
θ_{f_1}
θ_{f_2}
+
0.15
0.25
0.85
0.75
*
*
*
*
$\theta_{d_1|f_1b_1}$
$\theta_{d_1|f_1b_2}$
θ_{b_1}
θ_{b_2}
$\theta_{d_1|f_2b_1}$
$\theta_{d_1|f_2b_2}$
0.99
0.95
0.97
0.99
0.01
0.05
0.99
0.95
0.9
1
0.3
0.5



max = 0.7469
0.7543
0.15
0.25
0.85
0.75
0.99
0.95
0.97
0.99
0.01
0.05
0.99
0.95
0.9
1
0.3
0.5



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

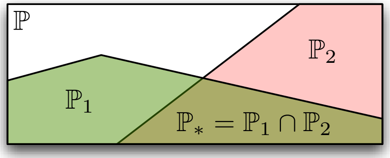

max = 0.7543
0.7543
$\overline{P}(f_2|d_1) = 0.7543$
$\underline{P}(f_2|d_1) = 0.4811$
$\theta_{d_1|f_1b_1}$
0.99
0.95
$\theta_{d_1|f_1b_2}$
0.97
0.99
0.01
0.05
0.99
0.95
$\theta_{d_1|f_2b_1}$
0.9
1
$\theta_{d_1|f_2b_2}$
0.3
0.5

4 Standard Semantics

General Idea

- Each premise $\varphi_i^{X_i}$ in $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^?$ is interpreted as a constraint $P(\varphi_i) \in X_i$ on the unknown prob. measure $P \in \mathbb{P}$

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



- The combined constraints of the premises may be

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

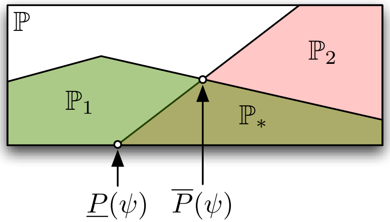
- General (under-determined) case: $Y = \{P(\psi) : P \in \mathbb{P}_*\}$

- 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

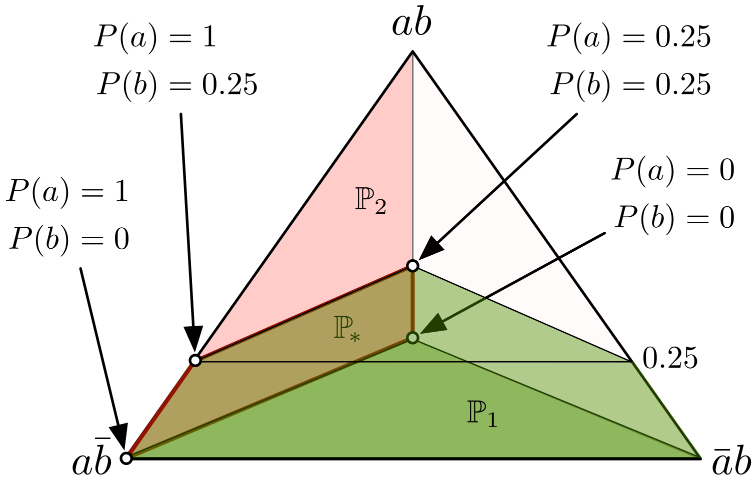
- If all probability sets X_i are intervals then
  - all sets $\mathbb{P}_i$ are convex
  - $\mathbb{P}_*$ is also convex
  - Y is also an interval, i.e. $Y = [\underline{P}(\psi), \overline{P}(\psi)]$, where $\underline{P}$ and $\overline{P}$ are vertices of $\mathbb{P}$

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



Example

- 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.



Inference

- normally a very large linear programming problem
- but if the premisses include independencies, credal nets become useful:
  - Construct the graph using e.g., Pearl’s modification of the PC algorithm
    - start with complete undirected graph
    - remove edges to capture independencies
    - orient the remaining edges

5 Probabilistic Argumentation

Given a question

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \mathrel{|\!\approx} \psi^?$$

separate the φ_i into uncertain versus certain

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_m^{X_m}, \varphi_{m+1}^{\{1\}}, \ldots, \varphi_n^{\{1\}} \mathrel{|\!\approx} \psi^?$$

Probabilistic Variables: $W =$ the set of propositional variables in $\varphi_1, \ldots, \varphi_m$.

States: Ω_W is the set of states of W.

Arguments: $Args(\psi) = \{\omega \in \Omega_W : \omega \wedge \varphi_{m+1} \wedge \cdots \wedge \varphi_n \models \psi\}$.

Conflicts: $Args(\bot)$, arguments inconsistent with certainties $\varphi_{m+1}, \ldots, \varphi_n$.

Non-Conflicts: $E = \Omega_W \backslash Args(\bot)$

Degree of Support: for P on W,

$$dsp_P(\psi) = P(Args(\psi)|E) = \frac{P(Args(\psi)) - P(Args(\bot))}{1 - P(Args(\bot))}$$

- Set $Y = \{dsp_P(\psi) : P \text{ satisfies } \varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_m^{X_m}\}$

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_m^{X_m}, \varphi_{m+1}^{\{1\}}, \ldots, \varphi_n^{\{1\}} \approx\!\!\!| \; \psi^?$$

N.B.

- $dsp_P(\psi) = P(Args(\psi)|E)$ is sub-additive as a function of ψ.
- If $m = 0$ we get classical logic.
- If $m = n$ we get the standard semantics.
- In general we get a non-monotonic logic.

Inference

$$Y = \{dsp(\psi) : dsp(\varphi_1) \in X_1, \ldots, dsp(\varphi_m) \in X_m\}$$

- Models are probability functions.
  - If the X_i are closed intervals then we can use credal nets.
    - Graph construction the same as with standard semantics.

6 Evidential Probability

First Order EP

- $\varphi_1, \ldots, \varphi_n$ are statements of a predicate language
  - these can include statistical statements of the form $freq_R(F) \in [l, u]$.
- given evidence $\varphi_1, \ldots, \varphi_n$, EP infers a statement ψ of the form $P(\theta) \in [l, u]$.

$$freq_R(F) \in [.2, .4], Rt \mathrel{|\!\approx} P(Ft) \in [.2, .4]$$

Second Order EP

- If the evidence is not perfectly reliable, how reliable is the conclusion?

$$freq_R(F) \in [.2, .4]^{[.9,1]}, Rt^{\{1\}} \mathrel{|\!\approx} P(Ft) \in [.2, .4]^{?}$$

Semantics

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \approx\!\!\!\mid\; \psi^Y$$

holds iff $P(\psi) \in Y$ for all probability functions P that satisfy

- the premisses $P(\varphi_1) \in X_1, \ldots, P(\varphi_n) \in X_n$,
- P is distributed uniformly over the EP interval, unless there is evidence otherwise,
- items of evidence are independent unless there is evidence of dependence.

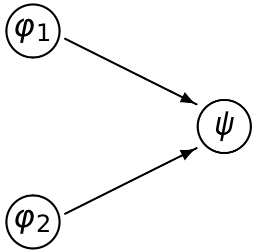
This yields a probabilistic logic (1^o EP is not)

Inference

- Models are probability functions.
  - If the X_i are closed intervals then we can use the EP semantics and credal nets.

The structure of 1^o EP calculations determine the structure of the credal net:

$$freq_R(F) \in [.2,.4]^{[.9,1]}, Rt^{\{1\}} \approx\!\!\models P(Ft) \in [.2,.4]^{?}$$



The X_i and 1^o EP inferences determine the conditional probability constraints:

$$P(\varphi_1) \in [.9,1], P(\varphi_2) = 1$$

$$P(\psi|\varphi_1 \wedge \varphi_2) = 1, P(\psi|\neg\varphi_1 \wedge \varphi_2) = .2 = P(\psi|\varphi_1 \wedge \neg\varphi_2) = P(\psi|\neg\varphi_1 \wedge \neg\varphi_2)$$

7 Bayesian Statistics

- Statistical hypotheses are themselves in the language: higher order probability.
- Uses Bayes' theorem to move from prior and likelihoods to posterior:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

- N.b. H typically determines $P(E|H)$.

Semantics

Given a question

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \mathrel{|}\!\approx \psi^?$$

separate the φ_i into uncertain versus certain

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_m^{X_m}, \varphi_{m+1}^{\{1\}}, \ldots, \varphi_n^{\{1\}} \mathrel{|}\!\approx \psi^?$$

- interpret the certain $\varphi_{m+1}, \ldots, \varphi_n$ as evidence E,
- interpret the uncertain $\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_m^{X_m}$ as information about the prior,
- suppose H imposes constraints $\chi_1, \ldots, \chi_k$ on the likelihoods,
  - then $Y = \{P(\psi|E) : P \text{ satisfies } \varphi_1^{X_1}, \ldots, \varphi_m^{X_m}, \chi_1, \ldots, \chi_k\}$.

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_m^{X_m}, \varphi_{m+1}^{\{1\}}, \ldots, \varphi_n^{\{1\}} \approx\!\!| \; \psi^{\{P(\psi|\varphi_{m+1},\ldots,\varphi_n):\varphi_1^{X_1},\ldots,\varphi_m^{X_m},\chi_1,\ldots,\chi_k\}}$$

Inference

- Models are probability functions.
- ▸ If the X_i are closed intervals then we can use credal nets.
- Build a credal net representing $\{P : P(\varphi_1) \in X_1, \ldots, P(\varphi_m) \in X_m, \chi_1, \ldots, \chi_k\}$
  - same methods as with standard semantics
- Update the credal net on evidence $\varphi_{m+1}, \ldots, \varphi_n$
- Use the common machinery for calculating $P(\psi)$ for each P represented by the net.
- ▸ Permits Bayesian inference without a fully specified prior.

8 Objective Bayesianism

An agent's degrees of belief should satisfy three norms:

Probability: they should be representable by a probability function $P \in \mathbb{P}$,

Calibration: they should be compatible with evidence $\mathcal{E}$: $P \in \mathbb{E} \subseteq \mathbb{P}$,

Equivocation: they should otherwise equivocate as far as possible

- i.e., they should be as close as possible to the equivocator $P_=$.

▸ they should be representable by $P_{\mathcal{E}} \in \downarrow\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E} : P \text{ is closest to } P_=\}$.

Semantics

Given a question

$$\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \mathrel{|\!\!\approx} \psi^?$$

- take the premisses to be evidence of empirical probability

$$\mathcal{E} = \{P^*(\varphi_1) \in X_1, \ldots, P^*(\varphi_n) \in X_n\}$$

$$\mathbb{E} = [\{P : P(\varphi_1) \in X_1, \ldots, P(\varphi_n) \in X_n\}]$$

- take Y to be $\{P(\psi) : P \in \downarrow\mathbb{E}\}$.

Inference

- Models are probability functions.

- If the X_i are closed intervals then we can use credal nets.
  - Create an undirected graph by linking all variables in the same constraint,
  - separation in this graph implies conditional independence for $P \in \downarrow\mathbb{E}$,
  - transform this graph into a dag satisfying the Markov condition,
  - determine the $P(A_i|Par_i)$ that maximise entropy,
  - use the common inferential machinery to determine $Y = \{P(\psi) : P \in \downarrow\mathbb{E}\}$.

Propositional Languages

- $\mathcal{L}$ has propositional variables $A_1, \ldots, A_n$
- An atomic state ω is a proposition of the form $\pm A_1 \wedge \cdots \wedge \pm A_n$

Probability: dobs should be representable by a probability function $P \in \mathbb{P}$,

1. $P(\omega) \geq 0$ for each $\omega \in \Omega$,
2. $\sum_{\omega \in \Omega} P(\omega) = 1$,
3. $P(\theta) = \sum_{\omega \models \theta} P(\omega)$ for each proposition θ.

Calibration: they should be compatible with evidence $\mathcal{E}$: $P \in \mathbb{E} \subseteq \mathbb{P}$,

Equivocation: they should be as close as possible to the equivocator $P_=$:

$$P_=(\omega) = \frac{1}{2^n}$$

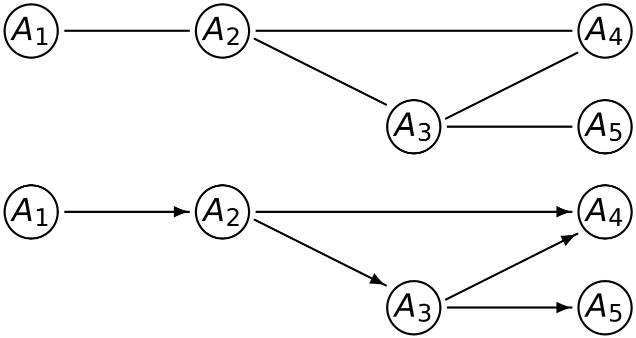
$$d(P, Q) = \sum_{\omega \in \Omega} P(\omega) \log \frac{P(\omega)}{Q(\omega)}$$

- ▸ dobs should be representable by $P_{\mathcal{E}} \in \downarrow\mathbb{E} = \{P \in \mathbb{E} : P \text{ minimises } d(P, P_=)\}$

Maximum Entropy Principle: An agent's degrees of belief should be representable by a probability function $P_{\mathcal{E}} \in \{P \in \mathbb{E} : H(P) \stackrel{\text{df}}{=} -\sum_{\omega \in \Omega} P(\omega) \log P(\omega) \text{ is maximised}\}$.

Example

$$A_1 \wedge \neg A_2^{[0.8,0.9]}, (\neg A_4 \vee A_3) \to a_2^{0.2}, A_5 \vee A_3^{[0.3,0.6]}, A_4^{0.7} \mid\!\approx A_5 \to A_1^{?}$$



$$\begin{aligned} P(A_5 \to A_1) &= P(\neg A_5 \wedge A_1) + P(A_5 \wedge A_1) + P(\neg A_5 \wedge \neg A_1) \\ &= P(A_1) + P(\neg A_5 | \neg A_1)(1 - P(A_1)) \end{aligned}$$

Summary

Framework. A unifying framework for probabilistic logic can be constructed around the fundamental question $\varphi_1^{X_1}, \varphi_2^{X_2}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!\models \psi^?$

Standard Semantics: $Y = \{P(\psi) : P$ satisfies premisses$\}$

Probabilistic Argumentation: Y = probability of worlds where entailment holds

Evidential Probability: Y = risk level associated with statistical inferences

Bayesian Statistics: Y = probabilities yielded by Bayes' theorem

Objective Bayesianism: $Y = \{P(\psi) : P$ represents beliefs on evidence of premisses$\}$

Calculus. Probabilistic networks can provide a calculus for probabilistic logic—they can be used to provide answers to the fundamental question

Network Construction: Build a credal net to represent those P that satisfy the premisses

Inference: Calculate Y from the net

A Reminder

Reading: Haenni, Romeijn, Wheeler, Williamson: *Probabilistic Logic and Probabilistic Networks*

Info: http://www.kent.ac.uk/secl/philosophy/jw/2006/progicnet/ESSLLI.htm

Schedule.

Day	Topic	Reading
Monday	The Progicnet Programme	§§1,8
Tuesday	Standard Semantics	§§2,9
Wednesday	Evidential Probability	§§4,11
	Probabilistic Argumentation	§§3,10
	Classical Statistics	§§5,12
Thursday	Bayesian Statistics	§§6,13
Friday	Objective Bayesian Epistemology	§§7,14