Probabilistic Logics and Probabilistic Networks
Lecture 5: Objective Bayesian Epistemology

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

Reading for this lecture:

- Probabilistic logics and probabilistic networks (Haenni et al., 2008) §§7,14.
- Objective Bayesian probabilistic logic (Williamson, 2008)
  - http://www.kent.ac.uk/secl/philosophy/jw/

1 Introduction

- Objective Bayesian epistemology (OBE): Jaynes (1957); Rosenkrantz (1977); Jaynes (2003) and Williamson (2005)

Maximum Entropy Principle: An agent's degrees of belief should be representable by a probability function, from all those that satisfy constraints imposed by her evidence, that has maximum entropy.

$$H(P) = -\sum_{\omega \in \Omega} P(\omega) \log P(\omega)$$

(Ω finite.)

- Nilsson (1986, §4): maxent could be used to facilitate inference in a progic.
- Here we show how OBE plugs into the progicnet programme:

  Representation: Questions for OBE can be represented by $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mathrel{|\!\!\approx} \psi^?$.

  Interpretation: $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mathrel{|\!\!\approx} \psi^Y$ interpreted using OBE.

  Inference: $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mathrel{|\!\!\approx} \psi^?$ answered using probabilistic nets.

  - Build a net representing the models of $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n}$,
  - Use progicnet machinery to determine $Y = \{P(\psi) : P \text{ models } \varphi_1^{X_1}, \ldots, \varphi_n^{X_n}\}$.

Contents

2 Objective Bayesian Epistemology

OBE addresses the following question:

- How strongly should an agent with evidence $\mathcal{E}$ believe the various propositions expressible in her language $\mathcal{L}$?

Language. We take $\mathcal{L}$ to be a propositional or predicate language.

Evidence. $\mathcal{E}$ includes everything she takes for granted in her current context.

- Background knowledge, observations, theoretical assumptions, etc.
- $\mathcal{E}$ may not be expressible in $\mathcal{L}$.

Propositional Languages

Three norms: Probability, Calibration, Equivocation.

Probability: The strengths of the agent's beliefs should be representable by probabilities.

- there is a probability function P on $\mathcal{L}$ such that $P(\theta)$ represents the degree to which the agent should believe θ, for each proposition θ expressible in $\mathcal{L}$.

  1. $P(\omega) \geq 0$ for each atomic state $\omega \in \Omega = \{\pm A_1 \wedge \cdots \wedge \pm A_n\}$,
  2. $\sum_{\omega \in \Omega} P(\omega) = 1$,
  3. $P(\theta) = \sum_{\omega \models \theta} P(\omega)$ for each proposition θ expressible in $\mathcal{L}$.

- $\mathbb{P}_{\mathcal{L}}$ = set of probability functions on $\mathcal{L}$.

- Why? degrees of belief are indicative of betting intentions,

  - these should be probabilities if the agent is to avoid bets that lose money whatever happens.

Calibration: The agent's degrees of belief should satisfy constraints imposed by $\mathcal{E}$.

- there is some set $\mathbb{E} \subseteq \mathbb{P}_{\mathcal{L}}$ compatible with $\mathcal{E}$ and $P_{\mathcal{E}} \in \mathbb{E}$.

- $\mathbb{E} = [\mathbb{P}^*] \cap \mathbb{S}$
  - $\mathcal{E}$ implies that empirical probability $P^* \in \mathbb{P}^* \Rightarrow P_{\mathcal{E}} \in [\mathbb{P}^*]$, non-empty convex closure.
  - $\mathbb{S}$: $P_{\mathcal{E}}$ should satisfy structural constraints imposed by qualitative evidence.
    - e.g., causal, logical, hierarchical or ontological structure.
    - not relevant in the probabilistic logic context.

- Why? Degrees of belief are used for inference and well-calibrated degrees of belief lead to more reliable inferences in the long run.

Equivocation: The agent's degrees of belief should otherwise be as equivocal as possible.

- i.e., as close as possible to the equivocator $P_=$ on $\mathcal{L}$:

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

- 'Distance': *cross entropy* or *Kullback-Leibler divergence*:

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

- Why? belief is a basis for action:
  - more extreme degrees of belief tend to trigger high-risk actions;
  - equivocal degrees of belief are associated with lower risks;
  - the agent should only take on risk to the minimum extent warranted by evidence.

Summary. The agent's degrees of belief should be representable by $P_{\mathcal{E}} \in \downarrow\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E} : d(P, P_{=}) \text{ is minimised}\}$.

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

- $P_{\mathcal{E}}$ is uniquely determined.
  - Entropy is a strictly concave function and it is being maximised over a closed and convex set of probability functions $\mathbb{E} = [\mathbb{P}^*]$, so it has a unique maximum.

3 Predicate Languages

- $\mathcal{L}$ is a first-order predicate language without equality.
- Each individual is picked out by a unique constant symbol t_i;
- Countable infinity $t_1, t_2, \ldots$ of such constants.
- Finitely many predicate symbols.
- For $n \geq 1$ let $\mathcal{L}_n$ be the finite predicate language on $t_1, \ldots, t_n$.
- Let $A_1, A_2, \ldots$ run through the atomic propositions of $\mathcal{L}$,
  - i.e., propositions of the form Ut.
  - Any atomic proposition expressible in $\mathcal{L}_n$ but not expressible in $\mathcal{L}_m$ for $m < n$ should occur later in the ordering than those atomic propositions expressible in $\mathcal{L}_m$.
- Let $A_1, \ldots, A_{r_n}$ be the atomic propositions expressible in $\mathcal{L}_n$.
- *Atomic n-states* $\Omega_n = \{\pm A_1 \wedge \cdots \wedge \pm A_{r_n}\}$.

Probability: The strengths of the agent's beliefs should be representable by probabilities.

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 each quantifier-free proposition θ, $P(\theta) = \sum_{\omega_n \models \theta} P(\omega_n)$ where n is large enough that $\mathcal{L}_n$ contains all the atomic propositions occurring in θ,
4. quantified statements are assigned probabilities via

$$P(\forall x \theta(x)) = \lim_{m \to \infty} P\left(\bigwedge_{i=1}^{m} \theta(t_i) \right)$$

$$P(\exists x \theta(x)) = \lim_{m \to \infty} P\left(\bigvee_{i=1}^{m} \theta(t_i) \right).$$

- P is determined by its values on the ω_n.

Calibration: The agent's degrees of belief should satisfy constraints imposed by her evidence.

- $P_{\mathcal{E}} \in \mathbb{E} = [\mathbb{P}^*] \cap \mathbb{S} = [\mathbb{P}^*]$ in the context of probabilistic logic.

Equivocation: The agent's degrees of belief should otherwise be as equivocal as possible.

- Define the equivocator $P_=$ on $\mathcal{L}$ by

$$P_=(\omega_n) = \frac{1}{2^{r_n}}$$

for all ω_n.

- Consider the n-distance

$$d_n(P,Q) = \sum_{\omega_n \in \Omega_n} P(\omega_n) \log \frac{P(\omega_n)}{Q(\omega_n)},$$

- P is *closer* to R than Q if there is some N such that for all $n \geq N$, $d_n(P,R) < d_n(Q,R)$.
  - Write $P \prec Q$ if P is closer to the equivocator $P_=$ than Q.
- Define $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E} : P \text{ is minimal with respect to } \prec\}$ as long as this set is non-empty
  - $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \mathbb{E}$ otherwise.

Summary. The agent's degrees of belief should be representable by $P_{\mathcal{E}} \in \downarrow\mathbb{E}$.

- Define the n-entropy $H_n(P)$ by

$$H_n(P) = -\sum_{\omega_n \in \Omega_n} P(\omega_n) \log P(\omega_n).$$

- Say that P has *greater entropy* than Q, written $P \gg Q$, if there is some N such that for all $n \geq N$, $H_n(P) > H_n(Q)$.

Maximum Entropy Principle: An agent's degrees of belief should be representable by a probability function $P_{\mathcal{E}} \in \{P \in \mathbb{E} : P \text{ is maximal with respect to } \gg\}$.

Discussion

Properties of the Closer Relation.

Proposition 3.1 *For fixed R the binary relation* $\cdot$ is closer than $\cdot$ to R *is irreflexive, asymmetric and transitive.*

Proposition 3.2 *If P is closer than Q to R then any proper convex combination of P and Q, i.e., $S = \lambda P + (1-\lambda)Q$ for $\lambda \in (0,1)$, is closer than Q to R.*

Example 3.3 *Suppose $\mathbb{E} = \{P : P(\forall x Ux) = c\}$ for some fixed $c \in [0,1]$. Define P by*

$$P(Ut_1) = \frac{c+1}{2}$$

$$P(Ut_{i+1} | Ut_1 \wedge \cdots \wedge Ut_i) = \frac{(2^{i+1}-1)c+1}{(2^{i+1}-2)c+2}$$

$$P(Ut_{i+1} | \pm Ut_1 \wedge \cdots \wedge \pm Ut_i) = \frac{1}{2}$$

otherwise. Then P is the member of $\mathbb{E}$ that is closest to the equivocator.

Definition of the ↓ Operator.

- $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E} : P$ is minimal with respect to $\prec\}$ if this set is non-empty, and $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \mathbb{E}$ otherwise.
- Are there any cases in which $\mathbb{E}$ has no minimal elements?
  - I.e., are there infinite descending chains wrt $\prec$?

Define P_i by

$$P_i(A_j|\omega_{j-1}) = \begin{cases} \frac{1}{2} & : \quad j < i \\ 1 & : \quad \text{otherwise} \end{cases}$$

for all j and all $\omega_{j-1} \in \Omega_{j-1}$.

- Then we have an infinite descending chain: for all i, $P_{i+1} \prec P_i$.

- But it is a normative constraint that $\mathbb{E}$ be closed.
  - Does $\{P_1, P_2, ...\}$ contains its limit points?
  - Depends on the notion of limit point.

Definition 3.4 (Strong Limit Point) *Probability function P is a* strong limit point *of $\mathbb{E}$ if for all $\varepsilon > 0$ there is some $Q \neq P$ in $\mathbb{E}$ such that $d_n(Q, P) < \varepsilon$ for all n.*

Definition 3.5 (Weak Limit Point) *Probability function P is a* weak limit point *of $\mathbb{E}$ if for all $\varepsilon > 0$ and for all n there is some $Q \neq P$ in $\mathbb{E}$ such that $d_n(Q, P) < \varepsilon$.*

- $\{P_1, P_2, \ldots\}$ has no strong limit points,
    - but $P_=$ is a weak limit point.
    - weak limit required.
    - $\downarrow\mathbb{E} = \{P_=\}$.

- Does every infinite descending chain have a weak limit point that is closer to the equivocator than any member of the chain?
- No.

Define Q_i by

$$Q_i(A_j|\omega_{j-1}) = \begin{cases} \frac{1}{2} & : \quad i < j \leq 2i \\ 1 & : \quad \text{otherwise} \end{cases}$$

for all j, ω_{j-1}.

- This defines an infinite descending chain: for all i, $Q_{i+1} \prec Q_i$.
- The only weak limit point is Q defined by $Q(A_j|\omega_{j-1}) = 1$ for all j, ω_{j-1}.
- But Q is a maximal element of $\mathbb{E}$: $Q_i \prec Q$ for all i.
  - So $\mathbb{E}$ has no minimal elements.

- take $\downarrow\mathbb{E} = \mathbb{E}$,
  - otherwise, if $\downarrow\mathbb{E}$ were a strict subset $\mathbb{F}$ of $\mathbb{E}$ in cases with infinite descending chains then idempotence could fail, $\downarrow\downarrow\mathbb{E} \neq \downarrow\mathbb{E}$.
- Equivocation becomes redundant.
- same conclusions as *Subjective Bayesian epistemology* (Probability and Calibration).

Equidistance.

Definition 3.6 (Equidistant) *P and Q are* equidistant *from R if neither is closer than the other to R.*

Proposition 3.7 *For fixed R the binary relation* equidistant *is reflexive and symmetric but not transitive in general and so not an equivalence relation.*

Definition 3.8 (Stably Equidistant) *P and Q are* stably equidistant *from R iff there is some N such that for all* $n \geq N$, $d_n(P,R) = d_n(Q,R)$.

Proposition 3.9 *For fixed R,* stably equidistant *defines an equivalence relation. If distinct P and Q are stably equidistant from R then any proper convex combination S of P and Q is closer than either P or Q to R.*

Definition 3.10 (Unstably Equidistant) *P and Q are* unstably equidistant *from R if they are equidistant from R but not stably equidistant from R.*

Proposition 3.11 *For fixed R,* unstably equidistant *is irreflexive, symmetric but not transitive in general. P and Q being unstably equidistant from R does* not *imply that any proper convex combination S of P and Q is closer than either P or Q to R. But a proper convex combination S of P and Q can be no further from R than P or Q. Nor can S and P (respectively S and Q) be stably equidistant from R.*

- $\downarrow\mathbb{E}$ is a (closed) convex set of probability functions.
  - Not necessarily a singleton with a predicate language.

4 Probabilistic Logics

Kinds of progic.

Internal: $\mathcal{L}$ includes function symbols that are interpreted as probability functions.

External: probabilities are attached to the propositions of $\mathcal{L}$.

Mixed: probabilities internal and external to $\mathcal{L}$.

- We will consider a mixed progic:

$$\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mathrel{|\!\!\approx} \psi^Y. \tag{1}$$

- $\varphi_1, \ldots, \varphi_n, \psi$ are propositions of $\mathcal{L}$,
- $X_1, \ldots, X_n, Y$ are sets of probabilities that attach to these propositions
- $\mathrel{|\!\!\approx}$ is an unspecified entailment relation.
- Propositional language $\mathcal{L}^\sharp$ with propositional variables of the form φ^X.
  - Language of the entailment relation $\mathrel{|\!\!\approx}$.

$$\mu_1, \ldots, \mu_n \mathrel{|\!\!\approx} \nu \tag{2}$$

- $\mu_1, \ldots, \mu_n, \nu$ are propositions of $\mathcal{L}^\sharp$.

Inferential Question.

$$\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models\!\!\approx \psi^? \tag{3}$$

More generally:

$$\mu_1, \ldots, \mu_n \models\!\!\approx \psi^? \tag{4}$$

- $\mu_1, \ldots, \mu_n$ are arbitrary propositions of $\mathcal{L}^\sharp$ and as before ψ is a proposition of $\mathcal{L}$.

5 Objective Bayesian Semantics

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

- Interpret $\mathcal{L}$ as the language of an agent,
- Interpret LHS as characterising her evidence $\mathcal{E}$:
  - φ^X is interpreted as $P^*(\varphi) \in X$, where P^* is empirical probability.
    - ▸ $\mathbb{P}^* = \{P \in \mathbb{P}_{\mathcal{L}} : P(\varphi_1) \in X_1, \ldots, P(\varphi_n) \in X_n\}$
    - ▸ $\mathbb{E} = [\mathbb{P}^*]$.
- Interpret RHS as a statement about the agent's rational degrees of belief: $P_{\mathcal{E}}(\psi) \in Y$.
- Interpret the relation as holding if that statement is forced to be true given the LHS:
  - $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mathrel{|\!\!\approx} \psi^Y$ iff any agent with evidence characterised by LHS should have degrees of belief satisfying the RHS.
  - iff $\{P(\psi) : P \in \downarrow\!\mathbb{E}\} \subseteq Y$.
- Similarly $\mu_1, \ldots, \mu_n \mathrel{|\!\!\approx} \nu$ iff any agent with $\mathcal{E}$ characterised by LHS has $P_{\mathcal{E}}$ satisfying RHS.

Fundamental Question. $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!| \; \psi^?$

- Interpreted as: to what extent should an agent with evidence of LHS believe ψ?
  - Similarly for $\mu_1, \ldots, \mu_n \approx\!\!\!\!| \; \psi^?$.
- Note that this is just the question that OBE asks:
  - OBE $\Longleftrightarrow$ progic fundamental question.
- Answer: $Y = \{P(\psi) : P \in \downarrow\mathbb{E}\}$.
  - Special case: LHS *regular* iff $\mathbb{P}^*$ is closed, convex and non-empty.
    - e.g., if mutually consistent, quantifier-free, with closed convex sets of probabilities.
  - ▸ $Y = \{P_{\mathcal{E}}(\psi) : P_{\mathcal{E}} \in \downarrow\mathbb{P}^*\}$.

Example 5.1 $\forall x U x^{3/5} \approx\!\!\!\!| \; Ut_1^?$*. The premiss is regular, so* $\mathbb{E} = [\mathbb{P}^*] = \mathbb{P}^* = \{P : P(\forall x Ux) = 3/5\}$*. There is one function P in* $\mathbb{E}$ *that is closest to the equivocator, as described in Example 3.3. This function gives* $P(Ut_i) = 4/5$ *for each constant* t_i*. Hence the answer to the question is* $4/5$*.*

- In the remainder of the paper we shall ignore infinite descending chains.

Discussion

- $\approx$ is a genuine semantic entailment relation: there are notions of *model* and *satisfies* such that $\mu_1, \ldots, \mu_n$ entails ν if and only if every model of $\mu_1, \ldots, \mu_n$ satisfies ν.

Satisfies: For proposition μ of $\mathcal{L}^\sharp$ let $\mathbb{P}(\mu) \stackrel{\text{df}}{=} \{P \in \mathbb{P}_\mathcal{L} : \mu \text{ holds}\}$.

Models: $\mathbb{M}(\mu) \stackrel{\text{df}}{=} \downarrow[\mathbb{P}(\mu)]$.

- Under OBE semantics $\mu_1, \ldots, \mu_n \approx \nu$ iff every model of the left-hand side satisfies the right-hand side,
  - i.e., iff $\mathbb{M}(\mu_1, \ldots, \mu_n) \subseteq \mathbb{P}(\nu)$.

Definition 5.2 (Decomposable) *Let* $\approx\!\!\!\!\mid$ *be an arbitrary entailment relation and let* $\mathbb{M}(\mu)$ *be the corresponding notion of a set of models of* μ. $\mid\!\approx$ *is called* decomposable *iff for all propositions* $\mu_1, \ldots, \mu_n$ *of the domain of the entailment relation,* $\mathbb{M}(\mu_1, \ldots, \mu_n) = \mathbb{M}(\mu_1) \cap \cdots \cap \mathbb{M}(\mu_n)$.

Definition 5.3 (Monotonic) *An entailment relation* $\mid\!\approx$ *is* monotonic *iff for all propositions* $\nu, \mu_1, \ldots, \mu_m, \ldots, \mu_n$ *of the domain of* $\mid\!\approx$ *(where* $m < n$*), we have that* $\mu_1, \ldots, \mu_m \mid\!\approx \nu$ *implies* $\mu_1, \ldots, \mu_m, \ldots, \mu_n \mid\!\approx \nu$.

Proposition 5.4 *A decomposable entailment relation is monotonic.*

- entailment under the objective Bayesian semantics is nonmonotonic:
  - $A_1^{[0,1]} \mid\!\approx A_1^{\{0.5\}}$ but it is not the case that $A_1^{[0,1]}, A_1^{\{1\}} \mid\!\approx A_1^{\{0.5\}}$.
- ▸ Hence objective Bayesian entailment is not decomposable.

System P Properties.

- A probability function P on $\mathcal{L}$ can be construed as a valuation on $\mathcal{L}^{\sharp}$:
  - P assigns the value True to proposition μ of $\mathcal{L}^{\sharp}$ if $P \in \mathbb{P}(\mu)$, i.e., if P satisfies μ.
- Define a decomposable entailment relation $\models$ by $\mathbb{M}_{\models}(\mu) = \mathbb{P}(\mu)$ for proposition μ of $\mathcal{L}^{\sharp}$.
  - In particular, if μ is of the form φ^X, then $P \in \mathbb{M}_{\models}(\mu)$ iff $P(\varphi) \in X$.
- $(\mathbb{P}_{\mathcal{L}}, \prec, \models)$ is a *preferential model*:
  - $\mathbb{P}_{\mathcal{L}}$ is a set of valuations on $\mathcal{L}^{\sharp}$,
  - $\prec$ is an irreflexive, transitive relation over $\mathbb{P}_{\mathcal{L}}$,
  - $\models$ is a decomposable entailment relation.
- Moreover, this preferential model is *smooth*:
  - if $P \in \mathbb{M}_{\models}(\mu)$ then either P is minimal with respect to $\prec$ in $\mathbb{P}(\mu)$ or there is a $Q \prec P$ in $\mathbb{P}(\mu)$ that is minimal.

▸ Hence this model determines a *preferential consequence relation* $\mid\!\sim$:

  - $\mu \mid\!\sim \nu$ iff P satisfies ν for every $P \in \mathbb{P}_{\mathcal{L}}$ that is minimal among those probability functions that satisfy μ.

- Wherever $\{\mu, \nu\}$ is regular, $\mid\!\sim$ will agree with $\mid\!\approx$.
- Consequently on regular propositions $\mid\!\approx$ will satisfy the properties of preferential consequence relations, often called *system-P* properties—see, e.g., Kraus et al. (1990):

Proposition 5.5 (Properties of Entailment) *Let $\models$ denote entailment in classical logic and let $\equiv$ denote classical logical equivalence. Whenever $\{\mu, \nu, \xi\}$ is regular,*

Right Weakening: *if $\mu \mid\!\approx \nu$ and $\nu \models \xi$ then $\mu \mid\!\approx \xi$.*

Left Classical Equivalence: *if $\mu \mid\!\approx \nu$ and $\mu \equiv \xi$ then $\xi \mid\!\approx \nu$.*

Cautious Monotony: *if $\mu \mid\!\approx \nu$ and $\mu \mid\!\approx \xi$ then $\mu \wedge \xi \mid\!\approx \nu$.*

Premiss Disjunction: *if $\mu \mid\!\approx \nu$ and $\xi \mid\!\approx \nu$ then $\mu \vee \xi \mid\!\approx \nu$.*

Conclusion Conjunction: *if $\mu \mid\!\approx \nu$ and $\mu \mid\!\approx \xi$ then $\mu \mid\!\approx \nu \wedge \xi$.*

6 Objective Bayesian Nets

We need some tools to answer questions of the fundamental form.

Bayesian nets

A *Bayesian net* is a potentially efficient representation of a probability function P.

DAG: On $A_1, \ldots, A_{r_n}$ propositional variables / atomic propositions of a finite propositional / finite predicate language $\mathcal{L}_n$.

Local Distributions: $P(A_i | Par_i)$.

Markov Condition: Each A_i is probabilistically independent of its non-descendants in the graph, conditional on its parents, written $A_i \perp\!\!\!\perp ND_i \,|\, Par_i$.

$$P(\omega_n) = \prod_{i=1}^{r_n} P(A_i^{\omega_n} | Par_i^{\omega_n})$$

Obnets

An *objective Bayesian net* is a Bayesian net representation of the rational belief function $P_{\mathcal{E}}$ on $\mathcal{L}_n$.

Construction.

1. determine independencies that must be satisfied by $P_{\mathcal{E}}$,
2. represent these independencies by a directed acyclic graph that satisfies the Markov Condition,
3. determine the conditional distributions $P_{\mathcal{E}}(A_i|Par_i)$.

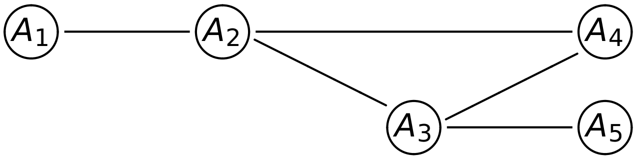
Step 1. Determine independencies that must be satisfied by $P_{\mathcal{E}}$.

Constraint graph: Take $A_1,\ldots,A_{r_n}$ as nodes and link two nodes with an edge if they occur in the same constraint in $\mathcal{E}$.

- If $\mathcal{L}_n$ is a predicate language and constraints involve quantifiers,
  - first substitute each occurrence of $\forall x\theta(x)$ by $\bigwedge_t \theta(t)$,
  - substitute each occurrence of $\exists x\theta(x)$ by $\bigvee_t \theta(t)$.

Proposition 6.1 *For all $P_{\mathcal{E}} \in \downarrow\mathbb{E}$, separation in the constraint graph implies conditional independence in $P_{\mathcal{E}}$: for $X,Y,Z \subseteq \{A_1,\ldots,A_n\}$, if Z separates X from Y in the constraint graph then $X \perp\!\!\!\perp Y \,|\, Z$ for $P_{\mathcal{E}}$.*

- E.g., $\mathcal{E}$: $P^*(A_1 \wedge \neg A_2) \in [0.8,0.9], P^*((\neg A_4 \vee A_3) \rightarrow A_2) = 0.2, P^*(A_5 \vee A_3) \in [0.3,0.6], P^*(A_4) = 0.7$.



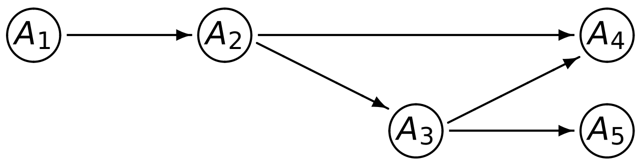
Step 2. Represent these independencies by a directed acyclic graph.

Algorithm 6.2

Input: An undirected graph $\mathcal{G}$.

- Triangulate $\mathcal{G}$ to give $\mathcal{G}^T$.
- Reorder the variables according to maximum cardinality search.
- Let $D_1, \ldots, D_l$ be the cliques of $\mathcal{G}^T$, ordered according to highest labelled vertex.
- Let $E_j = D_j \cap (\bigcup_{i=1}^{j-1} D_i)$ and $F_j = D_j \backslash E_j$, for $j = 1, \ldots, l$.
- Construct a directed acyclic graph $\mathcal{H}$ by taking propositional variables as nodes, and
  - Add an arrow from each vertex in E_j to each vertex in F_j, for $j = 1, \ldots, l$.
  - Add further arrows, from lower numbered variables to higher numbered variables, to ensure that there is an arrow between each pair of vertices in $D_j, j = 1, \ldots, l,$.

Output: A directed acyclic graph $\mathcal{H}$.



Step 3. Determine the conditional distributions $P(A_i|Par_i)$.
Use any of a variety of techniques, e.g.,

- numerical methods
- Lagrange multiplier methods.

$$H(P) = -\sum_{i=1}^{r_n}\sum_{\omega_i}\left(\prod_{A_j\in Anc_i} P(A_j^{\omega_i}|Par_j^{\omega_i})\right)\log P(A_i^{\omega_i}|Par_i^{\omega_i}),$$

- ▸ Finite language $\mathcal{L}_n$: unique entropy maximiser.

Infinite Languages.

- P determined by its values on the $\mathcal{L}_n$ for $n = 1, 2, \ldots$.
- ▸ Determined by the sequence of Bayesian nets representing P over the $\mathcal{L}_n$.
- Arrange that each net contains its predecessor in the sequence as a strict subnet.
- ▸ Define an *infinitary* Bayesian net to be the Bayesian net on the infinite domain $A_1, A_2, \ldots$ that has each member of the sequence of (finite) Bayesian nets as a subnet.
- ▸ $P_{\mathcal{E}} \in \downarrow\mathbb{E}$ may be represented by an infinitary obnet.

N.B.: $\downarrow\mathbb{E}$ may not be a singleton.

- All members of $\downarrow\mathbb{E}$ may be represented by (infinitary) obnets on the same graph (Proposition 6.1).

- $\downarrow\mathbb{E}$ is closed and convex: if $P, Q \in \downarrow\mathbb{E}$ then P, Q, S must be equidistant from the equivocator for any convex combination S of P and Q (since S can not be closer to the equivocator by definition of $\downarrow\mathbb{E}$, and neither can it be further from the equivocator by Proposition 3.11), so $S \in \downarrow\mathbb{E}$.

- ▸ $\downarrow\mathbb{E}$ can be represented by a *credal net*.
  - DAG + *contraints* on conditional probability distributions.
  - Equivalent to a set of Bayesian nets.

- An *objective credal net* (*ocnet*) is a credal net that represents $\downarrow\mathbb{E}$.

- Where $\mathcal{L}$ is an infinite predicate language, the ocnet will be infinitary.
  - Determine the dag as before via the constraint graph.

7 A Calculus for Probabilistic Logic

Recall

Question: $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\mid\!\!\!\approx \psi^?$

Semantics: OBE.

Answer: $Y = \{P_{\mathcal{E}}(\psi) : P_{\mathcal{E}} \in \downarrow[\mathbb{P}^*]\}$, where $\mathbb{P}^* = \{P : P(\varphi_1) \in X_1, \ldots, P(\varphi_n) \in X_n\}$. The question remains as to how to determine Y in practice.

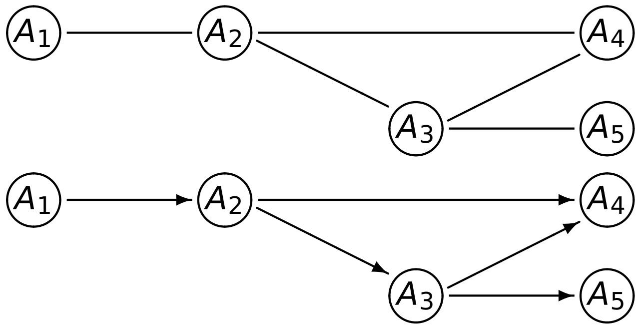
Progicnet programme: represent $\mathbb{E} = \downarrow[\mathbb{P}^*]$ by an (objective) credal net, and then use this net to calculate the probability interval Y that attaches to ψ.

Propositional Language Example

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

Interpretation: Supposing the agent's evidence says $P^*(A_1 \wedge \neg A_2) \in [0.8, 0.9]$, $P^*((\neg A_4 \vee A_3) \rightarrow A_2) = 0.2$, $P^*(A_5 \vee A_3) \in [0.3, 0.6]$, $P^*(A_4) = 0.7$, how strongly should she believe $A_5 \rightarrow A_1$?

- Construct an ocnet (= an obnet in this case).



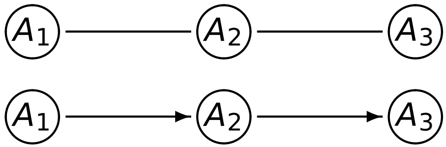
$$\begin{aligned} P(A_5 \rightarrow A_1) &= P(\neg A_5 \wedge A_1) + P(A_5 \wedge A_1) + P(\neg A_5 \wedge \neg A_1) \\ &= P(A_1) + P(\neg A_5 | \neg A_1)(1 - P(A_1)) \end{aligned}$$

- Calculate $P(A_1)$ and $P(\neg A_5 | \neg A_1)$ from the net.

Predicate Language Example

$$\forall x(Ux \to Vx)^{3/5}, \forall x(Vx \to Wx)^{3/4}, Ut_1^{[0.8,1]} \mid\!\approx Wt_1^{?}$$

- Only one constant symbol t_1, so we can focus on a finite predicate language $\mathcal{L}_1$.
- Let A_1 be Ut_1, A_2 be Vt_1 and A_3 be Wt_1.



- $P(A_1) = 4/5, P(A_2|A_1) = 3/4, P(A_2|\neg A_1) = 1/2, P(A_3|A_2) = 5/6, P(A_3|\neg A_2) = 1/2.$
  - Objective credal net (= objective Bayesian net).
  - Standard inference methods then give us the answer $P(A_3) = 11/15$.

General Procedure

Algorithm 7.1

Input: A question of the form $\mu_1, \ldots, \mu_n \mathrel{|\!\!\!\approx} \psi^?$, where $\mu_1, \ldots, \mu_n$ are propositions of $\mathcal{L}^\sharp$ and ψ is a proposition of $\mathcal{L}$.

1. If $\mathcal{L}$ is an infinite predicate language let n be the smallest j such that $\mu_1, \ldots, \mu_n$ are propositions of $\mathcal{L}_j^\sharp$ and ψ is a proposition of $\mathcal{L}_j$. Otherwise let $\mathcal{L}_n = \mathcal{L}$.
2. Construct the constraint graph $\mathcal{G}$ for constraints $\mu_1, \ldots, \mu_n$ on $\mathcal{L}_n$.
3. Transform this graph into a directed acyclic graph $\mathcal{H}$.
4. Determine the corresponding conditional probability intervals for an objective credal net representation of $\downarrow\mathbb{E}$:
   (a) Determine the credal net on $\mathcal{H}$ that represents the functions, from those that satisfy the Markov condition with respect to $\mathcal{H}$, that are in $\mathbb{E}$ (i.e., that satisfy the constraints $\mu_1, \ldots, \mu_n$ where these are consistent).
   (b) Use Monte-Carlo methods to narrow the intervals in this credal net to represent the probability functions, from those in the above net, that are closest to the equivocator.
5. Use this credal net to determine the interval Y that attaches to ψ.

Output: Approximate Y such that $\mu_1, \ldots, \mu_n \mathrel{|\!\!\!\approx} \psi^Y$.

Inference Step 5. Use the standard progicnet machinery.

Algorithm 7.2

Input: A credal net on $\mathcal{L}_n$ and a proposition ψ of $\mathcal{L}_n$.

- Compilation phase:
  - Transform the credal net into a d-DNNF (deterministic Decomposable Negation Normal Form) net.
- Inference phase:
  - Transform ψ into $\psi_1 \vee \cdots \vee \psi_l$ where the ψ_i are mutually exclusive conjunctions of literals—e.g., via Abraham's algorithm (Abraham, 1979).
  - Use hill-climbing in the d-DNNF net to calculate approximate bounds on each $P(\psi_i)$.
  - Calculate bounds on $P(\psi)$ via the identity $P(\psi) = \sum_{i=1}^{l} P(\psi_i)$.

Output: $Y = \{P(\psi) : P$ is subsumed by the input credal net$\}$.

8 Summary

OBE: Based on three norms: Probability, Calibration and Equivocation.

$$\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\models \psi^?$$

Representation: Standard OBE question take the form of the fundamental question.

Interpretation: Use OBE as a semantics for a question of this form.

Inference: Two step.

- Construct an objective credal net to represent the models of the LHS.
- Use standard machinery to calculate the range of probabilities to attach to ψ.

9 Course Summary

The Progicnet Programme

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

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

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

Inference: Calculate Y from the net

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

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

References

Abraham, J. A. (1979). An improved algorithm for network reliability. *IEEE Transactions on Reliability*, 28:58–61.

Haenni, R., Romeijn, J.-W., Wheeler, G., and Williamson, J. (2008). Probabilistic logic and probabilistic networks. *Forthcoming*. www.kent.ac.uk/secl/philosophy/jw/2006/progicnet.htm.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.

Jaynes, E. T. (2003). *Probability theory: the logic of science*. Cambridge University Press, Cambridge.

Kraus, S., Lehmann, D., and Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207.

Nilsson, N. J. (1986). Probabilistic logic. *Artificial Intelligence*, 28:71–87.

Rosenkrantz, R. D. (1977). *Inference, method and decision: towards a Bayesian philosophy of science*. Reidel, Dordrecht.

Williamson, J. (2005). *Bayesian nets and causality: philosophical and computational foundations*. Oxford University Press, Oxford.

Williamson, J. (2008). Objective Bayesian probabilistic logic. *Journal of Algorithms in Cognition, Informatics and Logic*, in press.