Probabilistic Logics and Probabilistic Networks
Lecture 3(a): Evidential Probability

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

ESSLLI 2008

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

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

1 Basic Theory

Calibration: probability assessments should be based upon relative frequencies, to the extent that we know them,

Total Evidence: the assignment of probability to specific events should be determined by everything that is known about that event.

- *Evidential certainties* Γ_δ.
  - Evidential propositions whose probability of error is less than δ.
- *Evidential probability* $Prob(\theta, \Gamma_\delta) = [l, u]$.
  - An interval rather than a sharp probability.
  - Interpret this as bounds on $P(\theta)$, the sharp probability of θ.
- *Practical certainties* Γ_ε are the inductive consequences of Γ_δ.
  - Propositions θ such that $Prob(\theta, \Gamma_\delta) \subseteq [1-\varepsilon, 1]$.
  - Risk of error is less than ε.

2 Kyburg's View

Kyburg was interested in the bounds, $Prob(\theta, \Gamma_\delta)$, rather than the sharp $P(\theta)$:

> we could say that any 'degree of belief' satisfying the probability bounds was 'rational'. But what would be the point of doing so? We agree with Ramsey that logic cannot determine a real-valued a priori degree of belief in pulling a black ball from an urn. This seems a case where degrees of belief are not appropriate. No particular degree of belief is defensible. We deny that there are any appropriate a priori degrees of belief, though there is a fine a priori probability: $[0, 1]$. There are real valued *bounds* on degrees of belief, determined by the logical structure of our evidence. (Kyburg Jr, 2003, p. 147)

So,

- evidence determines the interval $Prob(\theta, \Gamma_\delta)$,
- this interval can be thought of as bounding degree of belief, $P(\theta) \in Prob(\theta, \Gamma_\delta)$,
- logic does not determine a unique value for $P(\theta)$ within this interval.
  - N.B., Objective Bayesian epistemology holds that there are typically pragmatic reasons for choosing a particular value for $P(\theta)$ within the interval.
    - ▸ OBE can be viewed as an extension of EP.
    - ∗ See Friday's lectures.

3 Statistical Statements

$$\%x(\tau(x), \rho(x), [l, u]),$$

- the proportion of ρ-s that satisfy τ is between l and u: $freq_\rho(\tau) \in [l, u]$.
  - ρ is the *reference class*.
- The language $\mathcal{L}^{ep}$ of EP is a first order language in which such statements can be expressed.

Urn Example. It is known just that

- the proportion of White balls in an Urn is in $[l, u]$, %x(W(x), U(x), [l,u]).
- ball t is drawn from the Urn, $U(t)$.
- ▸ Then we can derive $P(W(t)) \in [l, u]$.
  - If $l = u = n/N$ then we can derive $P(W(t)) = n/N$.
  - If there is conflicting statistical evidence then EP applies conflict resolution rules.
  - If there is no statistical evidence we derive $P(W(t)) \in [0, 1]$.

4 Conflict Resolution

Reference Class Problem: How can one calculate the probability that an individual satisfies τ when it belongs to several reference classes ρ_i with known statistics?

- Suppose we know that
  - $[p, q]$ is the smallest interval covering reports of the proportion of Rs that are Us.
  - $[l, u]$ is the smallest interval covering reports of the proportion of Ss that are Vs.
  - $U(t_1) \longleftrightarrow V(t_2), R(t_1), S(t_2)$.
- $[p, q]$ and $[l, u]$ *conflict* if neither interval is strictly contained in the other.
- One can ignore the S, V evidence if

  Richness: The intervals conflict and R measures S together with other properties.
  - $R = (S, T, \ldots)$

  Specificity: The intervals conflict and R holds of fewer individuals than S.

  Strength: $[l, u]$ contains all intervals that survive Richness and Specificity.

- ▸ The remaining statistics are the *relevant statistics*. The smallest interval covering these gives the evidential probability.

Example

- Evidential certainties:
  - Bob smokes a packet of cigarettes a day and is a politician.

1. The proportion of smokers that live to the age of 80 is in $[.5, .8]$
2. The proportion of obese smokers living to 80 is in $[.3, .7]$ while the proportion of non-obese smokers living to 80 is in $[.6, .7]$.
3. The proportion of those who smoke a packet a day living to 80 is in $[.4, .75]$.
4. The proportion of politicians living to 80 is in $[.6, .7]$.

- 1 is eliminated in favour of 2 by Richness.
- 2 is eliminated in favour of 3 by Specificity.
- 3 is eliminated in favour of 4 by Strength.

▸ We conclude that the probability that Bob will live to 80 is in $[.6, .7]$.

5 Entailment

First-order EP. EP works like this:

- Evidence consists of statements $\varphi_1, \ldots, \varphi_n$ and their risk levels.
- Consider the evidential certainties, $\Gamma_\delta = \{\varphi_i : P(\varphi_i) \geq 1 - \delta\}$.
- From Γ_δ infer statements ψ of the form $P(\theta) \in [l, u]$.
- From these statements infer $\Gamma_\varepsilon = \{\theta : l \geq 1 - \varepsilon\}$ (the accepted conclusions).

$$\varphi_1^{[1-\delta,1]}, \ldots, \varphi_m^{[1-\delta,1]}, \varphi_{m+1}^{X_{m+1}}, \ldots, \varphi_n^{X_n} \longrightarrow \varphi_1, \ldots, \varphi_m \longrightarrow \psi \longrightarrow \theta$$

× This ignores

- the $\varphi_i \notin \Gamma_\delta$.
- the uncertainty attaching to each $\varphi_i \in \Gamma_\delta$.
- ▸ $[l, u]$ may not accurately bound the probability of θ given the evidence.

× This also ignores

- the $\theta \notin \Gamma_\varepsilon$
- the uncertainty of each $\theta \in \Gamma_\varepsilon$.

× Acceptance over a constant threshold is implausible.

<u>Second-order EP</u>. Total Evidence: it is better not to ignore the uncertainties:

- Evidence consists of $\varphi_1, \ldots \varphi_n$ and their risk levels $X_1, \ldots, X_n$.
- Infer statements ψ of the form $P(\theta) \in [l, u]$ *and their risk levels* Y.
- Decide what statements θ to accept on the basis of this fuller information.

Representation. $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\models \psi^Y$

Semantics. The entailment 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):
  - if $P(Ft) \in [l, u]$ follows from Φ by 1^o EP then

$$P(P(Ft) \in [l', u'] | \Phi) = \frac{|[l, u] \cap [l', u']|}{|[l, u]|},$$

- items of evidence are independent (unless there is evidence of dependence).

- ▸ Models are probability functions.
  - ▸ This yields a (decomposable, monotonic) probabilistic logic.
  - Not the case with 1^o EP.

6 Inference

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

- Suppose
  - LHS involves statistical statements,
  - ψ is of the form $P(\theta) \in [l, u]$.
- ▸ Natural to invoke the EP semantics.
  - Models are probability functions.
- If the X_i are closed intervals then we can represent the models of the LHS by a credal net.

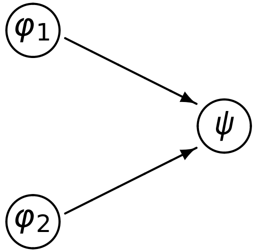
Plug and Play: Use this net to calculate $Y = \{P(\psi) : P \text{ models LHS}\}$

- Via the common inferential machinery:
  - Compilation methodology: compile to a d-DNNF net.
  - Use hill-climbing numerical methods to approximate Y.

Constructing the Credal Net

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:

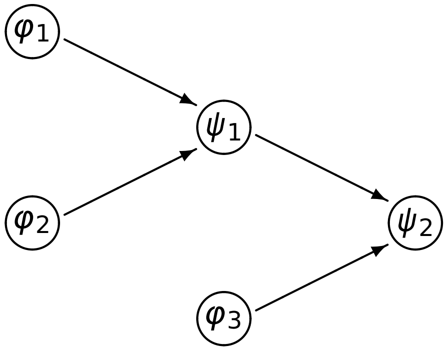
$$P(\varphi_1) \in [.9, 1], P(\varphi_2) = 1$$

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

We can then chain inferences (won't work in 1^o EP):

$$freq_R(F) \in [.2, .4]^{[.9,1]}, Rt^{\{1\}} \mid\!\approx P(Ft) \in [.2, .4]^{Y_1}$$

$$freq_R(F) \in [.2, .4]^{[.9,1]}, Rt^{\{1\}}, freq_F(G) \in [.6, .7]^{[.9,1]} \mid\!\approx P(Gt) \in [0, .25]^{?}$$



$$P(\varphi_1) \in [.9, 1], P(\varphi_2) = 1, P(\varphi_3) \in [0.9, 1]$$

$$P(\psi_1|\varphi_1 \wedge \varphi_2) = 1, P(\psi_1|\neg\varphi_1 \wedge \varphi_2) = .2 = P(\psi_1|\varphi_1 \wedge \neg\varphi_2) = P(\psi_1|\neg\varphi_1 \wedge \neg\varphi_2)$$

$$P(\psi_2|\psi_1 \wedge \varphi_3) = \frac{|[.2 \times .6 + .8 \times .1, .4 \times .7 + .6 \times .1] \cap [0, .25]|}{|[.2 \times .6 + .8 \times .1, .4 \times .7 + .6 \times .1]|} = .31,$$

$$P(\psi_2|\neg\psi_1 \wedge \varphi_3) = .27, P(\psi_2|\psi_1 \wedge \neg\varphi_3) = P(\psi_2|\neg\psi_1 \wedge \neg\varphi_3) = .25$$

References

Kyburg Jr, H. E. (2003). Are there degrees of belief? *Journal of Applied Logic*, 1:139–149.