Connections between Decision-making and Probabilistic Logic

Jon Williamson

University of Kent
CENTRE FOR REASONING
www.kent.ac.uk/reasoning

Progic2009, Groningen, 17-19 September

1 Logic and Decision-Making

What is the link between logic and decision-making?

None: Logic is about consequence / consistency, DM about pragmatics

✓ theoretical versus pragmatic rationality in traditional epistemology

Weak: A justificatory link between the two—e.g., a pragmatic justification of logic

✓ e.g., Dutch book considerations in Bayesian epistemology

Strong: Can't do one without the other, & how one is done depends on the other.

- DM requires logic and logic is determined by the DM context.

2 The Role of Probabilistic Logic

- Typically, one has, at any one time, probabilities of some propositions but not others.
- Decision theory requires probabilities of *relevant* propositions.
  - E.g., A utility matrix for judging chemotherapy:

		Judgement	
		C	$\neg C$
Case	R	5	-10
	$\neg R$	-4	1

- One needs a way of determining appropriate probabilities for relevant propositions from given probabilities.
- One needs probabilistic logic, which can answer questions of the form:

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

3 Probabilistic Logics

Expressions of the form $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!| \; \psi^Y$ admit a variety of semantics:

Standard Probabilistic 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

Bayesian Epistemology: Y = appropriate degree of belief in ψ

Probabilistic networks can provide a calculus for probabilistic logic—they can often 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

See:

- Rolf Haenni's progic2007 talk,
- Haenni, R., Romeijn, J.-W., Wheeler, G., and Williamson, J. (2010). *Probabilistic logic and probabilistic networks*. Synthese Library. Springer

4 Bayesian Epistemology

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

Y = appropriate degree of belief in ψ, given that $X_1, \ldots, X_n$ are appropriate for $\varphi_1, \ldots, \varphi_n$.

- I.e., $P(\psi) \in Y$ where $P(\varphi_1) \in X_1, \ldots, P(\varphi_n) \in X_n$ and P is a rational belief function.

4.1 Norms for Bayesian Epistemology

Probability: To avoid sure synchronic loss, P should be a probability function.

- ▸ We have a genuine probabilistic logic.

Calibration: To avoid sure expected or long-run loss, P should be calibrated with physical probability P^*, where known.

- Semantics: $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!| \; \psi^Y$ iff $P(\psi) \in Y$ where $P^*(\varphi_1) \in X_1, \ldots, P^*(\varphi_n) \in X_n$

Equivocation: To minimise worst-case expected loss, P should otherwise be closest to the equivocator $P_=$, where distance function d depends on the loss function.

- ▸ P is the *robust Bayes* choice
- if loss is logarithmic then d is KL-divergence and we get maxent.

Agent's language $\mathcal{L} = \{A_1, \ldots, A_n\}$, evidence $\mathcal{E}$, atomic states $\Omega = \{\pm A_1 \wedge \cdots \wedge \pm A_n\}$

Probability

P1: $P(\omega) \geq 0$ for each $\omega \in \Omega$,

P2: $P(\tau) = 1$ for some tautology $\tau \in S\mathcal{L}$, and

P3: $P(\theta) = \sum_{\omega \models \theta} P(\omega)$ for each $\theta \in S\mathcal{L}$.

Calibration

C: $P_{\mathcal{E}} \in \mathbb{E} = \langle \mathbb{P}^* \rangle \cap \mathbb{S}$

Equivocation

E: $P_{\mathcal{E}} \in \downarrow\mathbb{E} = \{P \in \mathbb{E} : d(P, P_=) \text{ is minimised}\}$

where $P_=(\omega) = 1/2^n$ for each $\omega \in \Omega$.

- But what is the distance function d?

(Williamson, J. (2010). *In defence of objective Bayesianism*. Oxford University Press, Oxford.)

4.2 Distance between probability functions

- $\mathcal{A}$: space of actions.
- $L(\omega, a)$: loss on doing $a \in \mathcal{A}$ when $\omega \in \Omega$ is the case.

Here $\mathcal{A} = \mathbb{P}$ and the loss function $L(\omega, Q)$ is called a *scoring rule*.

- $L(P, a) \stackrel{\text{df}}{=} E_P L(\Omega, a) = \sum_{\omega\in\Omega} P(\omega)L(\omega, a)$ is the *expected loss* for P.
- $H(P) \stackrel{\text{df}}{=} \inf_{a\in\mathcal{A}} L(P, a)$ is the *Bayes loss* or *generalised entropy* of P.

Assume that the scoring rule L is *proper*: for all P, $Q = P$ minimises $L(P, Q)$. (Q is a *Bayes act*.)

- ▸ Then $H(P) = L(P, P)$ for $P \in \mathbb{E}$.

- $d(P, Q) \stackrel{\text{df}}{=} L(P, Q) - H(P)$ is the *divergence* of P from Q.

Assume that the scoring rule L is *equivocator-neutral*: $L(P, P_=) = k$, a constant, for all P.

- e.g., brier score, logarithmic loss, zero-one loss.
- ▸ under natural conditions, $arg \inf_{Q\in\mathbb{P}} \sup_{P\in\mathbb{E}} L(P, Q) = arg \inf_{P\in\mathbb{E}} d(P, P_=)$
  - ▸ the functions minimising maximum expected loss are those in $\mathbb{E}$ closest to the equivocator.

Grünwald, P. and Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Annals of Statistics*, 32(4):1367–1433.

5 Logarithmic loss

- Log loss: $L(\omega, Q) = -\log Q(\omega)$
  - ▸ KL-divergence: $d(P, Q) = \sum_\omega P(\omega) \log \frac{P(\omega)}{Q(\omega)}$
  - ▸ Maxent: $P \in \mathbb{E}$ minimises $d(P, P_=)$ iff $P \in \mathbb{E}$ maximises $-\sum_\omega P(\omega) \log P(\omega)$
  - ▸ $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!| \; \psi^Y$ iff maxent P satisfying LHS satisfies RHS.
  - **?** Is log loss appropriate as a **default** loss function?

5.1 Information theory

- ✓ log as measure of information or code length.
  - × only relevant in particular cases.
    - ∗ E.g., loss = cost of communicating messages (Topsøe, 1979). Irrelevant here.
  - × N.b., Good (1952, §8) prefers a different logarithmic loss function.
- ✓ KL-divergence as a measure of distance
  - ✓ Hobson (1971): if $d(P, Q)$ is interpreted as the information in P that is not in Q.
  - ✓ Information geometry: divergence minimisation as projection.
    - × fits most naturally with exponential distributions.

5.2 Kelly Gambling

Betting set-up:

- a return of o_i pounds for each pound bet on ω_i if ω_i turns out true.
- agent bets $Q(\omega_i)W$ on each ω_i where W is her total wealth.
  - e.g., horse-racing works like this.
- bets are placed repeatedly.
- outcomes of the ω_i are assumed iid with respect to chance P^*,

▸ minimising divergence from the equivocator *maximises the worst-case expected growth rate* of W.

  × Rather particular to horse-race-like gambling scenarios.

  × Depends on betting one's total wealth on each race.

(Kelly, 1956; Cover and Thomas, 1991, Chapter 6; Grünwald, 2000, §5.1)

5.3 Properties of default loss

- By default, $L(\omega, Q) = 0$ if $Q(\omega) = 1$.
- By default, loss strictly increases as $Q(\omega)$ decreases from 1 towards 0.
- By default, loss $L(\omega, Q)$ depends on $Q(\omega)$ but not on $Q(\omega')$ for $\omega' \neq \omega$.
- By default, losses are presumed additive when domains are taken to be mutually irrelevant:
  - If $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ and $\mathcal{L}_1 \perp\!\!\!\perp_Q \mathcal{L}_2$ then $L_{\mathcal{L}}(\omega_1 \wedge \omega_2, Q) = L_{\mathcal{L}_1}(\omega_1, Q\restriction_{\mathcal{L}_1}) + L_{\mathcal{L}_2}(\omega_2, Q\restriction_{\mathcal{L}_2})$.

▸ Then loss is logarithmic, $L(\omega, Q) = -\log_b Q(\omega)$.

Contents

1 Logic and Decision-Making **2**

2 The Role of Probabilistic Logic **3**

3 Probabilistic Logics **4**

4 Bayesian Epistemology **5**

4.1 Norms for Bayesian Epistemology 5

4.2 Distance between probability functions 7

5 Logarithmic loss **8**

5.1 Information theory 9

5.2 Kelly Gambling 10

5.3 Properties of default loss 11

6 Judgement Aggregation **13**

7 Predicate Languages **17**

6 Judgement Aggregation

Difficulties

Discursive dilemma:

	θ	$\theta \rightarrow \varphi$	φ
A	true	true	true
B	true	false	false
C	false	true	false
Majority	true	true	false

Impossibility results (e.g., Dietrich and List, 2007): the only aggregation functions are *dictatorships* if the agenda is sufficiently rich and:

Universal Domain: the domain of the aggregation function is the set of all possible profiles of consistent and complete individual judgement sets,

Collective Rationality: the aggregation function generates consistent and complete collective judgement sets,

Independence: the aggregated judgement on each proposition depends only on individual judgements on that proposition,

Unanimity: if each individual judges a particular proposition true then so will the aggregate.

Judgement is a decision problem

A utility matrix for judging chemotherapy:

		Judgement	
		C	$\neg C$
Case	R	5	-10
	$\neg R$	-4	1

Decide in favour of chemotherapy if $EU(C) > EU(\neg C)$, i.e., if $P(R) > 1/4$.

Reasons for Judgements

Suppose agents $i = 1, \ldots, k$ give *reasons* $\varphi_{i_1}, \ldots, \varphi_{i_l}$ for their judgements concerning C: Then one can merge the reasons and ask about R, which determines the judgement:

$$\varphi_{i_1}^{X_1}, \ldots, \varphi_{i_l}^{X_1}, \ldots, \varphi_{k_1}^{X_k}, \ldots, \varphi_{k_l}^{X_k} \mathrel{|\!\!\approx} R^?$$

- Here X_i is an assessment of the reliability of agent i:
  - The probability that i is correct about φ_{i_j} is in X_i.

Suppose $\varphi_{i_1}^{X_1} \cup \cdots \cup \varphi_{k_l}^{X_k} \mathrel{|\!\!\approx} R^Y$.

- If $Y \subseteq [.25, 1]$ then judge C,
- If $Y \subseteq [0, .25]$ then judge $\neg C$,
- otherwise collect more evidence.

Note that decision-making is playing a dual role here:

- the specific decision problem determines the relation between $P(R)$ and the judgement on C,
- the general scoring rule determines the logic $\mathrel{|\!\!\approx}$.

Discursive dilemma again:

	θ	$\theta \to \varphi$	φ
A	true	true	true
B	true	false	false
C	false	true	false
Majority	true	true	false

Interpreting the middle two columns as the reasons:

$$\theta^{X_1}, \theta \to \varphi^{X_1}, \theta^{X_2}, \neg(\theta \to \varphi)^{X_2}, \neg\theta^{X_3}, \neg(\theta \to \varphi)^{X_3} \mathrel{|\!\!\approx} \varphi^{?}$$

Assume

- $X_1 = X_2 = X_3 = [0.5, 1]$,
- the threshold for judging φ / $\neg\varphi$ is 0.5,
- log loss scoring rule.

Then,

$$\theta^{X_1}, \theta \to \varphi^{X_1}, \theta^{X_2}, \neg(\theta \to \varphi)^{X_2}, \neg\theta^{X_3}, \neg(\theta \to \varphi)^{X_3} \mathrel{|\!\!\approx} \varphi^{0.25}$$

and an ‘aggregate agent’ (agent with aggregated reasons) should judge $\neg\varphi$.

- Goes against the majority view wrt reasons!
- Impossibility result does not apply since this method violates at least Independence and Unanimity.

7 Predicate Languages

- $\mathcal{L}$ is a first-order predicate language without equality.
- Each individual is picked out by a unique constant symbol t_i.
- Countably many constants $t_1, t_2, \ldots$
- Finitely many predicate symbols.
- $\mathcal{L}_n$ is the finite predicate language involving only $t_1, \ldots, t_n$.
- $A_1, A_2, \ldots, A_{r_n}$ are the atomic propositions of $\mathcal{L}_n$,
  - i.e., propositions of the form Ut.
- An *atomic n-state* ω_n is an atomic state $\pm A_1 \wedge \cdots \wedge \pm A_{r_n}$ of $\mathcal{L}_n$.
- Ω_n is the set of atomic n-states.

Probability

The agent's rational belief function is a function $P : S\mathcal{L} \longrightarrow \mathbb{R}$ that satisfies the properties

PP1: $P(\omega_n) \geq 0$ for each $\omega_n \in \Omega_n$ and each n,

PP2: $P(\tau) = 1$ for some tautology $\tau \in S\mathcal{L}$,

PP3: $P(\theta) = \sum_{\omega_n \models \theta} P(\omega_n)$ for each quantifier-free proposition θ, where n is large enogh that $\mathcal{L}_n$ contains all the atomic propositions occurring in θ, and

PP4: $P(\exists x \theta(x)) = \sup_m P\left(\bigvee_{i=1}^m \theta(t_i)\right)$.

Calibration

C: $P_{\mathcal{E}} \in \mathbb{E} = \langle \mathbb{P}^* \rangle \cap \mathbb{S}$

Equivocation

P should otherwise be sufficiently equivocal:

- Equivocator $P_=(\omega_n) = \frac{1}{2^{rn}}$ for all n, ω_n.
- n-distance: e.g., n-divergence $d_n(P,Q) = \sum_{\omega_n \in \Omega_n} P(\omega_n) \log \frac{P(\omega_n)}{Q(\omega_n)}$,
- 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 $\downarrow\mathbb{E}$ to be the set of members of $\mathbb{E}$ that are minimal with respect to $\prec$.
  - $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E} : \text{there is no } Q \in \mathbb{E} \text{ such that } Q \prec P\}$.

E: $P_{\mathcal{E}} \in \downarrow\mathbb{E}$.

e.g.,

$$\forall x Ux^{3/5} \mathrel{|\!\!\approx} Ut_1^{4/5}$$

$$\forall x(Vx \to Hx)^{[.6,1]}, \forall x(Hx \to Mx)^{[.75,1]}, Vs^{.8} \mathrel{|\!\!\approx} Ms^{11/15}$$

Papers. http://www.kent.ac.uk/secl/philosophy/jw

References

Cover, T. M. and Thomas, J. A. (1991). *Elements of information theory*. John Wiley and Sons, New York.

Dietrich, F. and List, C. (2007). Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, 29(1):19–33.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B*, 14(1):107–114.

Grünwald, P. (2000). Maximum entropy and the glasses you are looking through. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pages 238–246. Morgan Kaufmann.

Grünwald, P. and Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Annals of Statistics*, 32(4):1367–1433.

Haenni, R., Romeijn, J.-W., Wheeler, G., and Williamson, J. (2010). *Probabilistic logic and probabilistic networks*. Synthese Library. Springer.

Hobson, A. (1971). *Concepts in statistical mechanics*. Gordon and Breach, New York.

Kelly, J. (1956). A new interpretation of information rate. *Bell Systems Technical Journal*, 35:917–926.

Topsøe, F. (1979). Information theoretical optimization techniques. *Kybernetika*, 15:1–27.

Williamson, J. (2010). *In defence of objective Bayesianism*. Oxford University Press, Oxford.