1. Three desiderata for rational credence

COHERENCE Rational credences should be coherent
LEARNING Rational credences should allow learning on the basis of evidence
INVARIANCE Rational credences should be insensitive to irrelevant differences in the presentation of the epistemic situation

1.1. Coherence

TWO-SIDED BETTING

- A bet that wins $1 - x$ if A is true, and loses x if A is false is acceptable if $\mathbf{pr}(A) > x$.
- If the above bet is not acceptable then a bet that loses $1 - y$ if A and wins y otherwise is acceptable for any $y > x$.

Call a bet of this form – "win $1 - x$ if A, lose x otherwise" – a "unit bet on A (with betting quotient x)". The bet in the second clause here is a unit bet on $\neg A$ with betting quotient $1 - y$.

DE FINETTI COHERENCE

Your prices for gambles in a two-sided betting scenario should be such that no combination of acceptable bets is guaranteed to yield a loss in every state.

PROBABILITY FUNCTION

For a space of possibilities Ω, and a boolean algebra of propositions over Ω, $\mathcal{B}(\Omega)$, $\mathbf{pr} : \mathcal{B}(\Omega) \to [0, 1]$ is a probability function if:

- $\mathbf{pr}(\Omega) = 1$
- If X, Y are mutually exclusive then $\mathbf{pr}(X \vee Y) \geq \mathbf{pr}(X) + \mathbf{pr}(Y)$
- If X, Y are mutually exclusive then $\mathbf{pr}(X \vee Y) \leq \mathbf{pr}(X) + \mathbf{pr}(Y)$

Your betting quotients are de Finetti-coherent iff they are derived from a probability function (de Finetti 1964).

1.2. Learning

Let $\mathbf{pr}$ and $\mathbf{pr}'$ be your credences before and after learning that E is true and nothing else. Beyond requiring that $\mathbf{pr}$ and $\mathbf{pr}'$ are separately coherent, we should want them to be *jointly coherent*. This means:

CONDITIONALISATION

$\mathbf{pr}'(X) = \mathbf{pr}(X|E) = \frac{\mathbf{pr}(XE)}{\mathbf{pr}(E)}$

LEARNING ABOUT COIN FLIPS

- $H_1, H_2, \ldots$ are random variables encoding the outcome of tosses of a mystery coin
- $H_i = 1$ if the coin landed heads, 0 otherwise
- Let $S_n = \sum H_i$ be the number of heads in n tosses
- The event of interest is that the next coin toss lands heads H
- Define a probability function by $\mathbf{pr}(H) = \frac{\mu}{\mu+\nu}$ for $\mu, \nu > 0$
- And $\mathbf{pr}(H|S_n = h) = \frac{\mu+h}{\mu+\nu+n}$

Call these $\mathbf{pr}$ beta distributions.[1] By the law of large numbers, as n increases the ratio $\frac{S_n}{n}$ will tend to the true chance, and as n gets larger, $\frac{\mu+S_n}{\mu+\nu+n}$ gets closer to $\frac{S_n}{n}$. So each beta $\mathbf{pr}$ can learn. This is a good situation to be in.

Let E_n be the total evidence up to time n. And let $\mathbf{pr}^*$ be the "perfect" credence, whatever that is (the chance function?).

ALMOST SURE CORRECTNESS IN THE LIMIT

With probability 1, in the limit as $n \to \infty$, $\mathbf{pr}(H|E_n) = \mathbf{pr}^*(H)$

Note that this correctness is relative to a particular kind of proposition to be learned: in this case, the event of the next toss landing heads.

[1] Strictly speaking $\mathbf{pr}$ is the expectation of a beta distribution.

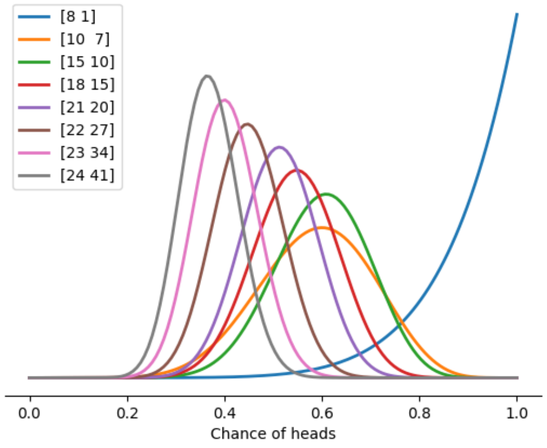

Figure 1: A beta distribution learns

1.3. Invariance

Consider a further kind of principle, "Principles of Indifference": unless you have a good reason for treating them differently, you should treat any two propositions the same.

> **Transformations**
>
> A transformation is a function $T : \Omega \to \Omega$ that induces a transformation on $\mathcal{B}(\Omega)$ through $T(A) = \{T(\omega) : \omega \in A\}$ for all $A \in \mathcal{B}(\Omega)$, and induces a transformation on the probability functions defined over $\mathcal{B}(\Omega)$ by $T(\mathbf{pr})(A) = \mathbf{pr}(T(A))$.

> **Precise invariance**
>
> A probability function $\mathbf{pr}$ is invariant under a collection of transformations $\mathcal{T}$ iff $T(\mathbf{pr}(A)) = \mathbf{pr}(A)$ for all $A \in \mathcal{B}(\Omega)$ and $T \in \mathcal{T}$.

Explicit invariance You have positive reason to treat A and $t(A)$ the same

Norm invariance In the absence of evidence to the contrary, the norms will treat A and $t(A)$ the same

Default invariance In the absence of evidence to the contrary, you ought to treat A and $t(A)$ the same

Call invariance with respect to the set of all transformations "universal precise invariance".

> **Impossibility I**
>
> No de Finetti-coherent credence satisfies universal precise invariance.

Possible responses:

- Go permissivist: deny default invariance, accept that you can arbitrarily break (norm invariance) symmetries
- Weaken invariance: (Paris and Vencovská 2015; Williamson 2010)
- Weaken coherence: this talk

2. Imprecise Probabilities

> **Credal sets and lower probabilities**
>
> Define:
>
> - $\mathbb{P}$, a set of prior probabilities with a common algebra of events. We shall call this a *credal set*.
> - $\mathbb{P}(X) = \{\mathbf{pr}(X), \mathbf{pr} \in \mathbb{P}\}$, the set-valued function defined over the same algebra as the members of $\mathbb{P}$.
> - $\overline{\mathbb{P}}(X) = \sup \mathbb{P}(X)$, the "upper probability".
> - $\underline{\mathbb{P}}(X) = \inf \mathbb{P}(X)$, the "lower probability".

2.1. Walley coherence

Is $\underline{\mathbb{P}}$ de Finetti coherent? No: Since $\underline{\mathbb{P}}$ is only superadditive, you might have $\underline{\mathbb{P}}(A) = \underline{\mathbb{P}}(\neg A) \leq x < y < \frac{1}{2}$. But if you're forced to accept the other sides of both bets for any price greater than x, you've paid $2(1 - y) > 1$ for two bets that guarantee you winnings of 1.

One-sided betting

- A bet that wins $1 - x$ if A is true, and loses x if A is false is acceptable if $\mathbf{pr}(A) > x$.

And on the basis of this constraint, we define some coherence requirements.

Avoid Sure Loss

Your prices for gambles in a one-sided betting scenario should be such that no combination of acceptable bets is guaranteed to yield a loss in every state.

Walley coherence

Your prices for gambles in a one-sided betting scenario should be such that no combination of acceptable bets is guaranteed to yield a loss in every state and every combination of bets that yield a gain in every state is acceptable.

Your betting quotients are Walley coherent iff they are the lower probability for some credal set (Troffaes and de Cooman 2014, Theorem 4.38).

2.2. Invariance weak and strong

A transformation of Ω induces a transformation on a credal set pointwise: $T(\mathbb{P}) = \{T(\mathbf{pr}) : \mathbf{pr} \in \mathbb{P}\}$. This allows us to make a distinction between two kinds of invariance (de Cooman and Miranda 2007).

Weak Invariance

$\mathbb{P}$ is weakly $\mathcal{T}$-invariant iff $T(\mathbb{P}) \subseteq \mathbb{P}$ for all $T \in \mathcal{T}, \mathbf{pr} \in \mathbb{P}$.

Strong Invariance

$\mathbb{P}$ is strongly $\mathcal{T}$-invariant iff $T(\mathbf{pr}) = \mathbf{pr}$ for all $T \in \mathcal{T}, \mathbf{pr} \in \mathbb{P}$.

As above, call (weak,strong) invariance with respect to the set of all transformations "universal (weak, strong) invariance".

Impossibility II

No Walley-coherent credence function satisfies universal strong invariance.

Vacuous prior

Let $\mathbb{V}$ be the set of all probability functions over the algebra of events.

Possibility I

The only Walley-coherent credence function that satisfies universal weak invariance is the *vacuous prior*.

2.3. Belief Inertia

Define a conditional credal set: $\mathbb{P}(-|E) = \{\mathbf{pr}(-|E), \mathbf{pr} \in \mathbb{P}, \mathbf{pr}(E) > 0\}$, the set obtained by conditioning each member of $\mathbb{P}$ on the same evidence. The vacuous prior $\mathbb{V}$ cannot learn.

Anti-inductive priors

- $\mathbf{pr}(h_1 \wedge h_2) = \mathbf{pr}(t_1 \wedge t_2) = \frac{\epsilon}{2}$
- $\mathbf{pr}(h_1 \wedge t_2) = \mathbf{pr}(t_1 \wedge h_2) = \frac{1-\epsilon}{2}$
- $\mathbf{pr}(h_1) = \mathbf{pr}(h_2) = \frac{1}{2}$
- $\mathbf{pr}(h_2|h_1) = \epsilon$

If you're accommodating inductive sceptics, you can't expect to learn! So Universal Weak Invariance is too much invariance.

Let $\mathbb{B}$ be the set of all beta distributions. This set of priors isn't vacuous, but it is "near-ignorance".

Near ignorance

$\mathbb{P}$ is near ignorance for $\mathcal{H}$ if, for all $H \in \mathcal{H}$, $(0,1) \subseteq \mathbb{P}(H)$.

Suppose we've gathered evidence $S_n = h$. $\mathbb{B}(H|S_n = h) = (0,1)$ even though every **pr** $\in \mathbb{B}$ satisfies ASCL. So what property learning-related property is $\mathbb{B}$ failing to satisfy?

INFORMATIVENESS FOR H

With probability one, $\overline{\mathbb{P}}(H|E_n) - \underline{\mathbb{P}}(H|E_n) \leq \delta_H(n)$ for some function δ_H such that for all $n > n_0$, $\delta_H(n) < 1$.

STRONG INFORMATIVENESS FOR H

With probability one, $\overline{\mathbb{P}}(H|E_n) - \underline{\mathbb{P}}(H|E_n) \leq \delta_H(n)$ for some function δ_H such that $\delta_H(n) \to 0$ as $n \to \infty$.

IMPOSSIBILITY III

No Walley-coherent credence function satisfies universal weak invariance and informativeness.

3. Learning by ignoring the stubborn

ALTERNATIVE PARAMETRISATION

- $\phi = \mu + \nu$
- $\mu' = \frac{\mu}{\phi}$ and $\nu' = \frac{\nu}{\phi}$

Therefore:

- $\mu' + \nu' = 1$
- $\mathbf{pr}(H) = \frac{\mu'}{\mu' + \nu'}$,
- ϕ governs how "quickly" the distribution learns

After learning $S_n = h$:

- $\phi_{\text{new}} = \phi + n$
- $\mu'_{\text{new}} = \frac{h + \phi\mu'}{n + \phi}$

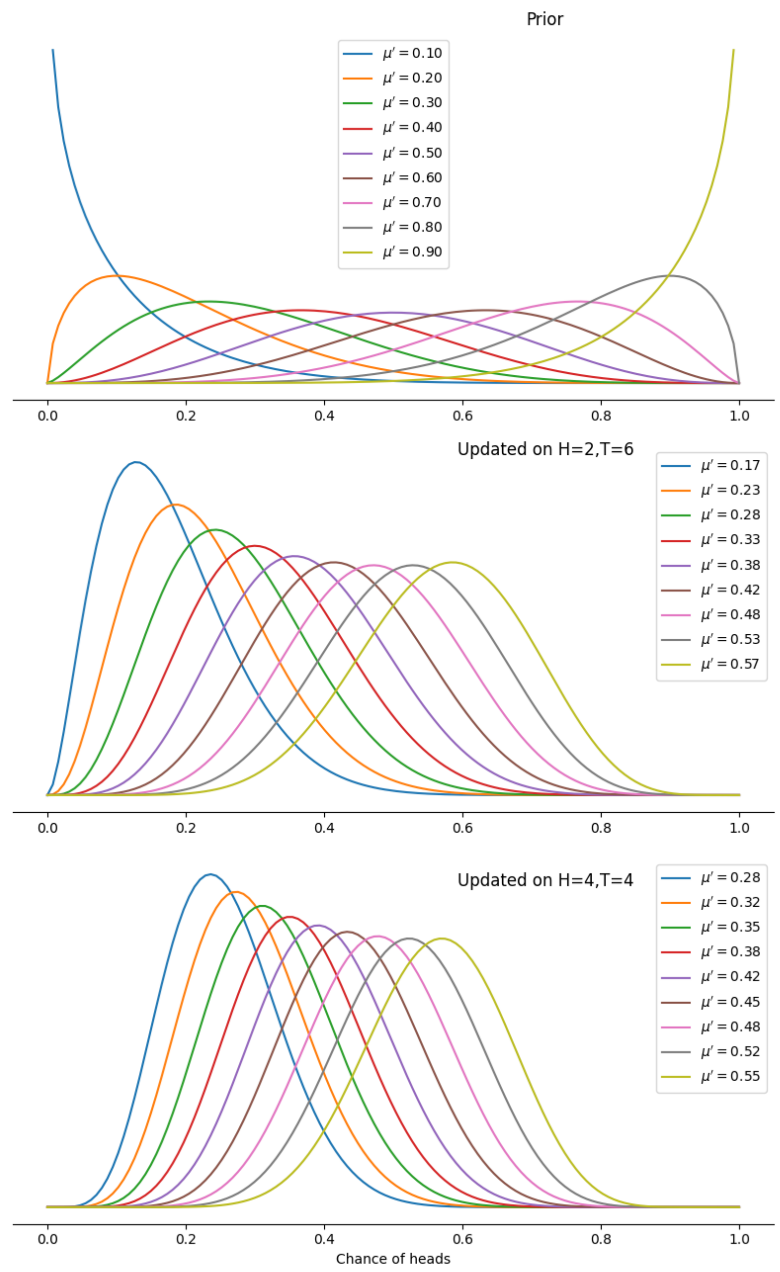

Figure 2: The alternative parametrisation, holding ϕ fixed

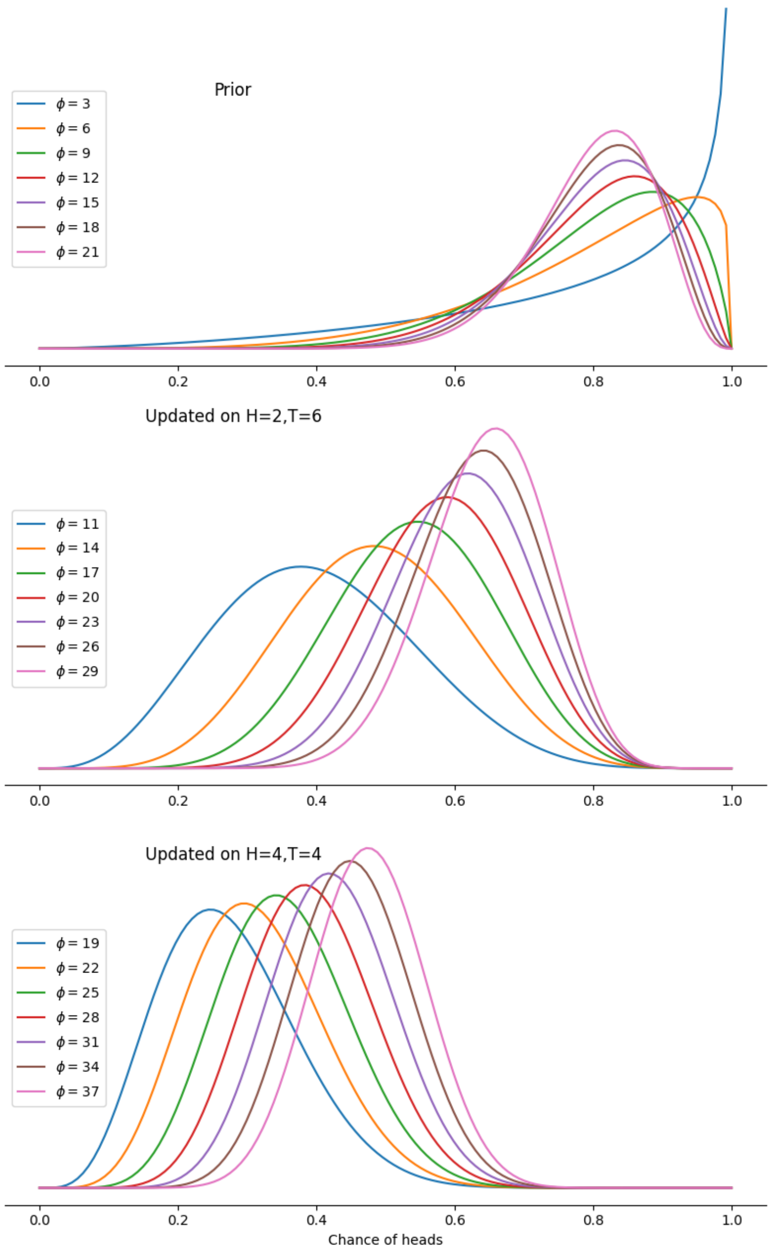

Figure 3: The alternative parametrisation, holding μ' fixed

Therefore:

$$\underline{\mathbb{B}}(H|S_n = h) = \frac{S_n}{n+\phi} \qquad \overline{\mathbb{B}}(H|S_n = h) = \frac{S_n + \phi}{n+\phi}$$

So solve belief inertia by setting an upper bound on ϕ. $\mathbb{B}_\phi$ satisfies strong informativeness for any $\phi < \infty$. Call this model IDM (Walley 1996).

Exchangeability

$\mathbb{P}$ should be strongly invariant to permutations of the indices for the coin-toss propositions $H_1, H_2, \ldots, H_n$.

Representation insensitivity

$\mathbb{P}$ should be weakly invariant to swapping each H_i with T_i, and also to splitting up of one of the categories.

Impossibility IV

No de Finetti-coherent credence satisfies Representation Insensitivity.

Possibility II

IDM priors satisfy Walley-coherence, Representation Insensitivity, Exchangeability and Strong Informativeness.

See de Cooman, De Bock, and Diniz (2015), Theorems 21 and 4.

4. Learning by ignoring the most wrong

IDM relies on the evidence (and the data-generating process) having a certain structure. Lots of inference problems don't have that structure.

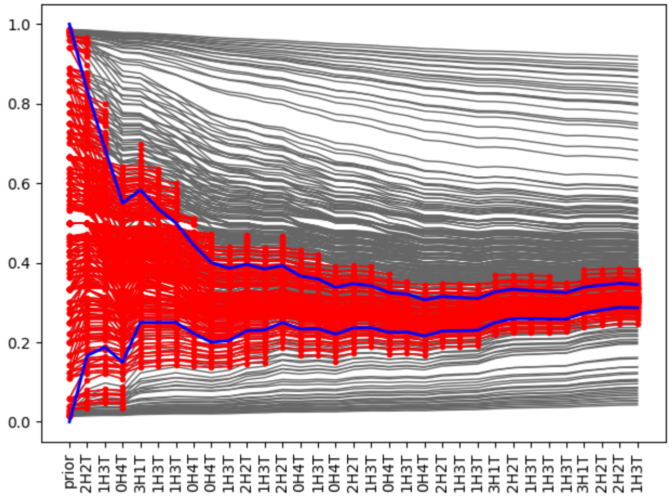

Figure 4: Alpha cut learning

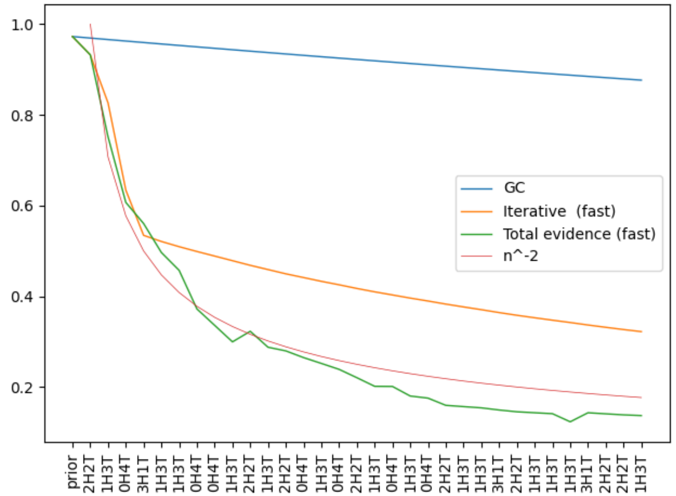

Figure 5: Spread of $\mathbf{pr}(H)$

α-CUT CONDITIONING

- $\mathbb{P}(X|_{\mathbb{Q}}E) = \{\mathbf{pr}(X|E); \mathbf{pr} \in \mathbb{Q} \subseteq \mathbb{P}\}$
- $\mathbb{Q} = \{\mathbf{pr} \in \mathbb{P}, \mathbf{pr}(E) \geq \alpha\overline{\mathbb{P}}(E)\}$, for $\alpha \in (0,1)$

In Figure 4 the grey lines represent GC;[2] the red lines α-cut; the blue lines are $\overline{\mathbb{B}}_\phi(H|S_n)$ and $\underline{\mathbb{B}}_\phi(H|S_n)$ for an IDM prior. If we graph $\overline{\mathbb{B}}(H|_\alpha S_n) - \underline{\mathbb{B}}(H|_\alpha S_n)$ as n grows (Figure 5), we see that it appears to shrink at a rate of about $\frac{1}{\sqrt{n}}$. The prior satisfies the same invariance as IDM. The conditional model probably doesn't.

$\mathbb{B}$ and $\mathbb{B}(-|_\alpha E)$ are each separately coherent, however, sometimes they are not jointly coherent, nor even jointly avoid sure loss.

SURE LOSS FOR α-CUT

Lottery A is a fair lottery with n tickets. You know nothing about lottery B, except that it also has n tickets so you have a near-ignorance prior for outcomes from lottery B. I'm going to flip a fair coin to decide whether to draw a lottery A ticket (heads) or a lottery B ticket (tails). Before I flip the coin, I'll offer you a bet at even odds that the coin lands heads. I'm then going to tell you the outcome of the draw from the urn. I then offer you a bet against heads at worse than even odds. (Pay $1 - \frac{1}{2n\alpha}$ to win 1 if tails). Unless α is small enough (i.e. less than $\frac{1}{n}$), these bets lose you money, whichever outcome I announce.[3]

Can you always avoid sure loss by making α small enough? Open question, but my guess is that you probably can for finite partitions at least. It would also be interesting to explore *how* incoherent α-cut conditioning is (Schervish, Seidenfeld, and Kadane 1997; Staffel 2020).

5. Conclusion

A rough summary:

ARBITRARY PRECISE PRIOR Excellent coherence, Mixed (?) learning, Poor invariance

PRECISE BETA PRIOR Excellent coherence, Good learning, Poor invariance

[2] i.e. $\mathbb{Q} = \{\mathbf{pr}(E) > 0\}$

[3] Thanks to Marco Cattaneo for this example.

PURE INDUCTIVE LOGIC Excellent coherence, Good learning, Moderate invariance

VACUOUS PRIOR Good coherence, Zero learning, Excellent invariance

IDM Good coherence, Good learning, Good invariance

α-CUT Moderate coherence, Mixed (?) learning, Mixed (?) invariance

It's striking how difficult it is to do justice to these three kinds of desiderata for rational credence. This paper contributes to recognising this deep and enduring tension in mathematical models of rational belief and inference.

A. Iterating α-cut

- $\mathbb{Q}_\alpha[E, \mathbb{P}] = \{\mathbf{pr} \in \mathbb{P}, \mathbf{pr}(E) \geq \alpha\overline{\mathbb{P}}(E)\}$
- $\mathbb{P}^\alpha_E = \mathbb{P}(-|_\alpha E) = \{\mathbf{pr}(-|E), \mathbf{pr} \in \mathbb{P}, \mathbf{pr}(E) \geq \alpha\overline{\mathbb{P}}(E)\}$

When you then learn F, after already having learned E what do you do?

- $\mathbb{P}(-|_\alpha EF)$ (total evidence update, down and right on Figure 6)
- $\mathbb{P}^\alpha_E(-|_\alpha F)$ (iterative update, down and left on Figure 6)

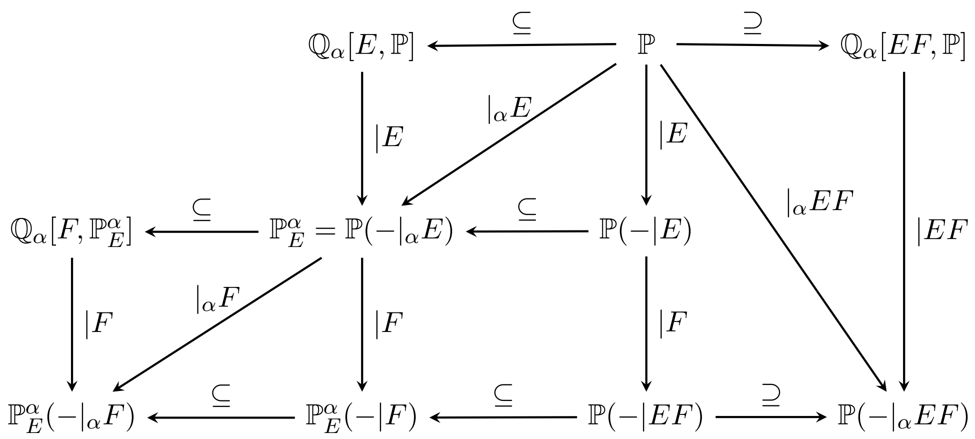

Figure 6: α-cut conditioning as GC plus restriction

References

Cattaneo, Marco (2014). "A Continuous Updating Rule for Imprecise Probabilities". In: *Information Processing and Management of Uncertainty in Knowledge Based Systems*. Ed. by Anne Laurent et al. Springer, pp. 426–435.

de Cooman, Gert, Jasper De Bock, and Márcio Alves Diniz (2015). "Coherent Predictive Inference under Exchangeability with Imprecise Probabilities". In: *Journal of Artificial Intelligence Research* 52, pp. 1–95.

de Cooman, Gert and Enrique Miranda (2007). "Symmetry of models versus models of symmetry". In: *Probability and Inference: Essays in Honor of Henry E. Kyburg Jnr.* Ed. by William Harper and Gregory Wheeler. Kings College Publications, pp. 67–149.

de Finetti, Bruno (1964). "Foresight: Its Logical Laws, Its Subjective Sources". In: *Studies in Subjective Probability*. Ed. by Henry E. Kyburg and Howard E. Smokler. Wiley, pp. 97–158.

Paris, Jeffrey and Alena Vencovská (2015). *Pure Inductive Logic*. Cambridge University Press.

Schervish, Mark J., Teddy Seidenfeld, and Joseph B. Kadane (1997). *Two Measures of Incoherence: How Not to Gamble If You Must*. Tech. rep. Department of Statistics, Carnegie Mellon University.

Staffel, Julia (2020). *Unsettled Thoughts*. Oxford University Press.

Troffaes, Matthias and Gert de Cooman (2014). *Lower Previsions*. Wiley.

Walley, Peter (1996). "Inferences from Multinomial Data: Learning about a Bag of Marbles". In: *Journal of the Royal Statistical Society B* 58.1, pp. 3–57.

Williamson, Jon (2010). *In Defense of Objective Bayesianism*. Oxford University Press.