Normality and Probability

Jeremy Goodman (USC) and Bernhard Salow (Oxford)

3 September 2021

Outline

Normality Structures

A *normality structure* is a tuple $\langle S, \mathcal{E}, W, \succcurlyeq, \ggcurly \rangle$ with:

1. S a non-empty set (of *states*),
2. $\mathcal{E} \subseteq \mathcal{P}(S) \backslash \{\varnothing\}$ (the *possible bodies of evidence*)
3. $W = \{\langle s, E \rangle : s \in E \in \mathcal{E}\}$ (the set of *(centered) worlds*),[1]
4. $\succcurlyeq$ a preorder on W (read as '*is at least as normal as*'),
5. $\ggcurly$ an asymmetric, well-founded relation on W (read as '*is sufficiently more normal than*'), such that,:
   5.1 If $w_1 \ggcurly w_2$, then $w_1 \succcurlyeq w_2$;
   5.2 If $w_1 \succcurlyeq w_2 \ggcurly w_3 \succcurlyeq w_4$, then $w_1 \ggcurly w_4$.

[1] Throughout, when I write $\langle s, E \rangle$, I assume that this is in W.

Belief

Your beliefs at $\langle s, E \rangle$ are all only those things true throughout $R_b(\langle s, E \rangle)$, defined as follows:

$$R_b(\langle s, E \rangle) := \{s' : s' \in E \& \neg(\exists s'' : \langle s'', E \rangle \gg \langle s', E \rangle)\}$$

'Believe that the state is not much less normal/plausible than the evidence requires it to be.'

Key innovation: the states left open are not just the most normal ones, but also all that are not sufficiently less normal than these.

Outline

Probability Structures

A *probability structure* is a tuple $\langle S, \mathcal{E}, W, Q, P, t\rangle$ with:

1. $S, \mathcal{E}, W$ as above,
2. Q (the *question*) a partition of S,
3. P (the *prior*) a probability distribution over S with $P_E(q) := P(q|E)$ defined for all $q \in Q$ and $E \in \mathcal{E}$,
4. $t \in (0, 1]$ (the *threshold*).

The Reduction, Part I

Let $[s]_Q :=$ the cell of Q containing s (the *answer* to Q at s).

NORMALITY AS LIKELINESS:

$\langle s, E\rangle \succcurlyeq \langle s', E\rangle := P_E([s]_Q) \geqslant P_E([s']_Q)$

The Reduction, Part II

Let $\tau(\langle s, E\rangle) := P_E(\{s' : \langle s, E\rangle \succcurlyeq \langle s', E\rangle\})$ – i.e., the probability, given E, that things are no more normal than they are at $\langle s, E\rangle$.

$$\text{SUFFICIENCY: } \langle s, E\rangle \ggcurly \langle s', E\rangle := 1 - \frac{\tau(\langle s', E\rangle)}{\tau(\langle s, E\rangle)} \geqslant t$$

Intuitively: conditional on things being no more normal than they are at $\langle s, E\rangle$, they are still $\geqslant t$ likely to be more normal than they are at $\langle s', E\rangle$.

Upshot

Context supplies a question (a partition of the set of worlds); the evidence produces a ranking of the complete answers to that question (the cells of the partition) by their probability.

Consider disjunctions of complete answers that respect this ranking, in that they contain any answer at least as probable as any other they contain.[2]

The idea: *the shortest such disjunction that has probability* $\geqslant t$ *characterizes your beliefs.*

[2] Compare Holguín (ms), Dorst & Mandelkern (forthcoming).

Outline

Bayesian Statistics

When dealing with continuous quantities, NORMALITY AS LIKELINESS should be replaced with the analogous principle about probability *densities*.[3] SUFFICIENCY can stay as it is.

The result: your beliefs are characterized by the 'highest posterior density region' — a way of summarizing a distribution popular in Bayesian statistics (as a rival to frequentist *confidence intervals*).[4]

[3] See Goodman & Salow (2021, Appendix B).
[4] See e.g. Hyndman (1996).

The Stability Theory of Belief

In some ways this is similar to Leitgeb's (2014) theory:

- Beliefs are consistent and closed under logical consequence.
- High probability is *necessary* for belief.
- Belief is sensitive to a question.

In other ways it is different:

- Leitgeb's theory, but not ours, maintains that high probability is *sufficient* for belief.
- Leitgeb's theory, but not ours, is consistent with strong classical dynamic principles (e.g. AGM).
- Our theory, but not Leitgeb's, allows for non-trivial beliefs even when the question is very fine-grained.

Outline

Flipping for Heads

Flipping for Heads: You flip a fair coin until it lands heads, observing the outcomes as they happen.

$S = \{1, 2, \dots\}$ (number of times coin will be flipped)

$\mathcal{E} = \{\{k : k \geqslant n\} : n \in S\}$

$P(\{n\}) = 2^{-n}$, $t = .99$, $Q = \{\{s\} : s \in S\}$

Predictions:

$\langle n, E \rangle \succcurlyeq \langle m, E \rangle$ iff $n \geqslant m$; and $\langle n, E \rangle \ggcurly \langle m, E \rangle$ iff $n \geqslant m + 7$

You always believe that the coin will land heads within the next 7 trials.

Racing for Heads I

Racing for Heads: Each of 10 coin flippers has a fair coin that they will flip until it lands heads. You wonder what will happen.

If $Q =$ *how many times will coin i be flipped*, then your beliefs about coin i will be like your beliefs about a single coin in **Flipping for Heads**. But you believe nothing about e.g.

- how the other coins will land,
- how many tails-landings there will be across all the coins,
- how many trials there will be before every coin has landed heads, or
- how many coins will ever land heads on the same flip.

Racing for Heads II

But you can have non-trivial beliefs about these topics relative to other natural values of Q:

Q	which worlds most normal	t	min tails	max tails	min trials	max trials	same end?
(i) exact outcome	all coins land heads first time	.75	0	13	1	14	?
		.95	0	18	1	19	?
(ii) outcome shape	6×1 flip, 3×2 flips, 1×3 flips	.75	1	15	2	8	no
		.95	0	22	1	12	?
(iii) how many total tails	8 or 9 total tails [tied]	.75	5	14	1	15	?
		.95	2	18	1	19	?
(iv) how long until over	ends on 4^{th} trial	.75	2	50	3	6	?
		.95	1	70	2	8	?
(v) how many end together	5 flippers get heads at once	.75	3	∞	2	∞	no
		.95	2	∞	2	∞	no

Outline

Modelling Dynamics

The result of *discovering* p in $\langle s, E \rangle$ is $\langle s, E \cap p \rangle$.

Preservation

PRESERVATION: If you believe p, and you don't believe not-q, you'll still believe p upon discovering q.[5]

This fails in **Flipping for Heads**:

- It's consistent with your beliefs that the coin will land tails the first time.
- Discovering that it did leads you to give up your belief that it won't take more than 7 flips to land heads.

[5] 'Rational Monotonicity' in Kraus et. al. (1990). Standard ordering accounts invalidate this if the normality order isn't total, which $\gg$ isn't.

Against Preservation

Counterexamples like this are very hard to deny:

- Deny that you have *any* beliefs about what will happen?
  - Then it's consistent with your beliefs that the coin lands tails the first 1,000,000 times.
  - But in any realistic scenario, you should stop believing the coin is fair when you discover that this happened.
  - So PRESERVATION still fails.
- Taking multiple measurements of a quantity with an imperfect measuring device (e.g. your weight with a scale), leads to similar dynamics.[6]

[6] Goodman and Salow (ms)

Learning from the Expected

NO LEARNING FROM THE EXPECTED: If you believe p, and you don't believe q, you still won't believe q upon discovering p.[7]

This fails in a variant of **Flipping for Heads**:

- Instead of observing the flips, you just discover (after the fact) whether there were more than seven flips.
- You already believe that there weren't.
- Upon discovering that there weren't, you form a new belief that there were no more than six.

A gloss: discovering p turns it from 'inductive' to 'evidential', freeing up inductive resources to support further beliefs.

[7] 'Cut' in Kraus et. al. (1990)

A Concern

Underlying this counterexample to NO LEARNING FROM THE EXPECTED is a failure of:

> E/$\gg$ INDEPENDENCE:
> $\langle s, E \rangle \gg \langle s', E \rangle$ if and only if $\langle s, E' \rangle \gg \langle s', E' \rangle$.

A general tension:

- E/$\gg$ INDEPENDENCE needs to fail if we want to reconcile substantial beliefs in cases like **Flipping for Heads** with a high probability requirement.
- But without E/$\gg$ INDEPENDENCE, dynamics look very unconstrained.

A Solution

But consider:

> E/$\succcurlyeq$ INDEPENDENCE:
> $\langle s, E\rangle \succcurlyeq \langle s', E\rangle$ if and only if $\langle s, E'\rangle \succcurlyeq \langle s', E'\rangle$

Unlike E/$\gg$ INDEPENDENCE, this holds whenever all discoveries about the state are *wholly about* Q, in the sense that they are unions of cells of Q. Often, this will be a natural assumption.[8]

[8] But not when e.g. allowing *de se* questions, which also raise intuitive problems for the principles below (Goodman & Salow 2021, Appendix C).

Principles

With E/$\succcurlyeq$ INDEPENDENCE, we get interesting principles that are not valid on simple threshold views:[9]

> NO REVERSAL: If you believe p and don't believe not-q, you won't believe not-p upon discovering q.
>
> PRESERVATION FROM THE EXPECTED: If you believe both p and q, you'll still believe p upon discovering q.[10]

We also get principles not generally valid in systems that allow non-total normality orders, e.g.

> ANTICIPATION: If you won't believe p upon discovering q, and you won't believe p upon discovering not-q, you already don't believe p.[11]

[9] Even their joint weakening NO REVERSAL FROM THE EXPECTED is only valid on simple threshold views if we require $t > .618$.

[10] 'Cautious Monotonicity' in Kraus et. al. (1990).

[11] 'Negation Rationality' in Kraus et. al. (1990).

Conclusion

Modelling belief with two closely related normality orders, $\succcurlyeq$ and $\ggcurly$, is really productive:

- It makes room for a straightforward connection between normality and probability.
- It gives the option of imposing strong constraints (totality, E-independence) on $\succcurlyeq$, which filter down into 'golden mean' intermediate-strength constraints on $\ggcurly$.
- The resulting dynamic principles are weaker than the (too strong) AGM principles, but sufficiently substantive to be worth studying.

References

- Dorst, K. and Mandelkern, M. (forthcoming) "Good Guesses" *Philosophy and Phenomenological Research*
- Goodman, J. and Salow, B. (2018) "Taking a Chance on KK" *Philosophical Studies* 175: 183-196
- Goodman, J. and Salow, B. (2021) "Knowledge from Probability" In J. Halpern and A. Perea (eds) *Proceedings of TARK 2021*
- Goodman, J. and Salow, B. (ms) "Epistemology Normalized"
- Holguín, B. (ms) "Thinking, Guessing, and Believing"
- Hyndman, R. (1996) "Computing and Graphing Highest Density Regions" *The American Statistician* 50: 120-126
- Kraus, S., Lehmann, D., and Magidor, M. (1990) "Nonmonotonic reasoning, preferential models, and cumulative logics" *Artificial Intelligence* 44: 167-207
- Leitgeb, H. (2014) "The Stability Theory of Belief" *Philosophical Review* 123: 131-171