

Remarks on Martin’s “False confidence, non-additive beliefs, and valid statistical inference”

Jesse Clifton

February 13, 2019

1 Non-additive beliefs are inaccurate

A different way to conceive of the reliable formation of beliefs in propositions — in this case, propositions like $\{\theta \in A\}$ — is as the *estimation* of their truth-values. This is the point of view taken in the literature on probabilistic forecasting, where a forecaster’s performance is often measured by their (empirical) *Brier score*, which is just the mean squared-error of the probabilities they assigned to past events with respect to the indicators of whether those events actually happened; see e.g. the work of Tetlock and colleagues (Mellers et al., 2015). It is the approach to justifying degrees of belief taken by many contemporary philosophers (Joyce, 2009; Pettigrew, 2016). And it is an application of the statistical principle that an estimator should be close, on average, to the true parameter — only here the “parameter” in question is just the truth-value of the hypothesis about which we are forming degrees of belief.

Call a generic proposition q ; then, its truth value is $1(q)$. On the accurate-estimation view, degrees of belief in q based on data Y — written $c(Y)$ — are justified if they are close (in some sense) on average (with respect to Y) to $1(q)$.

Non-additive beliefs are, in *this* sense, unreliable. We consider degrees of belief $c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ in pairs of propositions $Q = \begin{bmatrix} 1(q) \\ 1 - 1(q) \end{bmatrix}$, i.e. c_1 is the degree of belief that q and c_2 is the degree of belief that not- q . $Q \in \{e_1, e_2\}$ where e_1, e_2 are standard basis vectors. Then, define the *Brier score* as the squared error of c as an estimator of Q :

$$Brier(c, Q) \triangleq \|c - Q\|^2.$$

The following theorem says that degrees of belief b which are *not* probability distributions are Brier-dominated by a probability distribution.

Theorem 1. Let $Proj_\Delta$ be the projector onto the probability simplex Δ in \mathbb{R}^2 . Then for all $c \in \mathbb{R}^2$, and for $Q \in \{e_1, e_2\}$, where $\{e_1, e_2\}$ are standard basis vectors in \mathbb{R}^2 (representing worlds where q is true and false, respectively), $Brier(Proj_\Delta c, Q) < Brier(c, Q)$.

(Theorem 1 is a (constructive) special case of Joyce (1998)’s result that non-probabilistic degrees of belief are dominated by probabilistic ones in loss functions with certain desirable properties.)

This means that data-dependent degrees of belief over Q which are not probability distributions are not only inadmissible as estimators of Q , but are dominated everywhere by their projection onto Δ .

Corollary 1.1. Let $c(Y)$ be a function mapping data Y to degrees of belief, with $P(c(Y) \notin \Delta) = 1$. Then, for $Q \in \{e_1, e_2\}$, $Brier(Proj_\Delta c(Y), Q) < Brier(c(Y), Q)$ almost surely.

Returning to the case of forming beliefs about parameters in particular, this means that for a hypothesis of interest $H = \{\theta \in A\}$ and a non-additive belief-forming procedure $c(Y) = \begin{bmatrix} Bel_Y(\theta \in A) \\ Bel_Y(\theta \notin A) \end{bmatrix}$, the *additive* belief-forming procedure $Proj_\Delta c(Y)$ will always yield degrees of belief which are at least as close to the truth-value of H as $c(Y)$.

This is, of course, not a statement about degrees of belief over all of Θ (e.g. functions on the Borel σ -algebra on Θ), but only targeted hypotheses. Nevertheless this seems to be the setting

implicit in Martin’s discussion, and so the inaccuracy of belief functions with respect to such hypotheses seems relevant to evaluating his approach. I talk about this more in Section 2.

In any case, one is free to reject this competing notion of reliability. I only mean to point out that there are multiple senses in which a belief-forming procedure may systematically lead to error (relative to alternatives), and that non-additive degrees of belief are unreliable on one plausible account. I would be excited to see those interested in the foundations of statistics give more thorough consideration to exactly what sort of reliability we want from our methods.

2 Individual hypotheses versus beliefs over all of Θ

Martin motivates his rejection of additive degrees of belief with the *false confidence theorem*, which says that additive degrees of belief are miscalibrated for some hypotheses. One might respond that in the “inference” setting we are only interested in a few hypotheses of interest, and so that it should be enough that we avoid false confidence for those hypotheses. Martin acknowledges this, but says that “at the meta level where data scientists are developing *methods* and corresponding software to be used by others across applications, this strong control on performance is necessary”.

As far as I can tell this does not really answer the objection, because it does not address the possibility that we use methods which are reliable *given* a fixed (but arbitrarily chosen) hypothesis. This would allow for the development of generic, reliable methods which do not require anticipating problematic hypotheses. Indeed, as Martin discusses later, we already have a candidate for such a methodology in frequentist hypothesis testing, which does not involve forming beliefs (additive or otherwise) over the whole parameter space. Another approach is to identify procedures which, for fixed hypotheses $H = \{\theta \in A\}$, return accurate degrees of belief — in the sense of Section 1 — in $(H, \text{not-}H)$.

3 Continuous estimation and decision-making

A final concern I have on the general approach is that, as with frequentist hypothesis testing, it does not seem to lead to a satisfying account of continuous parameter estimation (as opposed to forming beliefs about coarse partitions of parameter space) or, as a consequence, decision-making based on a statistical analysis. Given a probability distribution over the decision-relevant observables Y (whether it be a posterior predictive distribution or otherwise), it is clear — if I accept that probability distribution as “justified” in some sense, and I accept the principle of maximizing expected utility — how to make predictions and decisions, at least in principle. Martin, of course, denies that any such probability distribution can be justified (because of false confidence), but it is not clear what the alternative should be in situations where one needs to quantify their uncertainty over Y “simultaneously”. It does seem that there have been several attempts to develop a decision theory for belief functions (Denoeux, 2018), but none seems to have emerged as a real competitor to expected utility theory.

4 Proof of theorem 1

Proof. Take $c \in \mathbb{R}^2$ and $Q \in \{e_1, e_2\}$. Proj_Δ is a projector onto a convex set, and therefore a strict quasi-nonexpansion. And, $Q \in \Delta$ and therefore Q is a fixed point of Proj_Δ . So, by the definition of strict quasi-nonexpansiveness, $\text{Brier}(\text{Proj}_\Delta c, Q) = \|\text{Proj}_\Delta c - Q\|^2 < \|c - Q\|^2 = \text{Brier}(c, Q)$. (See Bauschke et al. (2011) for the results on quasi-nonexpansions.) \square

References

- Heinz H Bauschke, Patrick L Combettes, et al. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.
- Thierry Denoeux. Decision-making with belief functions: a review. *arXiv preprint arXiv:1808.05322*, 2018.
- James M Joyce. A nonpragmatic vindication of probabilism. *Philosophy of science*, 65(4):575–603, 1998.

James M Joyce. Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In *Degrees of belief*, pages 263–297. Springer, 2009.

Barbara Mellers, Eric Stone, Terry Murray, Angela Minster, Nick Rohrbaugh, Michael Bishop, Eva Chen, Joshua Baker, Yuan Hou, Michael Horowitz, et al. Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science*, 10(3):267–281, 2015.

Richard Pettigrew. *Accuracy and the Laws of Credence*. Oxford University Press, 2016.