Comment on *False confidence, non-additive beliefs, and valid statistical inference* by Ryan Martin

Harry Crane*

March 12, 2019

1 *Primum non nocere*

The phrase is often attributed to the writings of Hippocrates and to the Hippocratic oath taken by physicians.[1] If I were to summarize Martin's perspective of statistical inference in one line, it would be this. *First, to do no harm.* This sentiment is implicit in Martin's emphasis on Reid and Cox's admonition to avoid "systematically misleading conclusions", advice echoed at least eight times by Martin and made precise by the proposed validity condition (Martin, 2019, Section 4).

Though I see a connection between the Reid–Cox mantra and the Statistician's Oath to do no harm, the two are not in full agreement. On the one hand, I can see how Martin's validity condition lives up to the Reid–Cox principle. On the other hand, I believe this condition to be insufficient for doing no harm. In fact, it seems desirable in some instances, the satellite conjunction analysis being one, to favor "systematically *wrong* conclusions" in the interest of doing no harm. I discuss further below.

1.1 Validity

Formally, the validity condition requires

$$\sup_{\theta \notin A} \mathbb{P}_{Y|\theta}(b_Y(A) > 1-\alpha) \leq \alpha, \quad 0 \leq \alpha \leq 1, \quad A \subseteq \Theta. \tag{1}$$

In words, as articulated by Martin (Martin, 2019, p. 13), (1) ensures that it is unlikely (under the assumed model) that the data will lead to high degree of belief in a false assertion. I take this condition as the guiding principle of Martin's approach, from which he derives the False Confidence Theorem (Theorem 1 in Martin, 2019) and proceeds to make the case for non-additive "degrees of belief" in his definition of an Inferential Model (IM). I focus primarily on the implications of this condition here.

2 Inferential Models

Martin's definition of an inferential model associates each possible observation $y \in \mathbb{Y}$ to a function $b_y : 2^\Theta \to [0,1]$, with the intended interpretation that "$b_y(A)$ represents the

*Rutgers University, Department of Statistics

[1] A little research into the origins suggests that this phrase does not appear in the Hippocratic oath.

data analyst's degree of belief in the truthfulness of the assertion '$\theta \in A$' ".

Because this definition is crucial to Martin's approach, I'll discuss some of its main components in detail.

2.1 The analyst's "degree of belief"

I suspect that Martin, and I assume many other statisticians and philosophers, is comfortable with the representation of "degrees of belief" about assertions A by a function that assigns a numerical value $b(A)$ to a set A corresponding to that assertion. For those comfortable with this setup, the concept of "degree of belief" may be regarded as sufficiently primitive that it requires no further discussion or explanation. To me, however, I find the concept nebulous and impossible to understand. I promise, I'm not just being difficult.

My own inadequacy aside, I want to point out two concerns with the use of this language that might hinder the IM framework's acceptance among those who understand—or think they understand—how functions b_y represent "degrees of belief".

First, and most obviously, the language of "degree of belief" has long been associated to the Bayesian subjectivist school of thought. For those ingrained in that mindset, and even many who aren't, this definition of IM might immediately suggest the Bayesian approach as the only viable alternative.

Second, and relatedly, those pushed toward the Bayesian mindset by this definition will be inclined to find arguments for why Definition 1 makes Bayesianism inevitable. One such path to this conclusion has been highlighted by a previous reviewer (Jesse Clifton), whose comments I'll discuss in the following section.

Before proceeding, I note that Martin goes to considerable lengths to differentiate his approach from the Bayesian one, and anyone who takes the time to understand the IM framework will easily see how it differs from Bayesianism. I wish to highlight this potential issue, however, because there are a number of directions one can go based on Definition 1. Martin goes in the direction of validity described in Section 1 above. A traditional Bayesian, in particular one subscribing to "accuracy-first epistemology", is inclined in another direction which, for the moment at least, is the more predominant point of view.

2.2 Accuracy First

As Clifton points out, non-additive beliefs are inaccurate in a precise sense that for any non-additive function b there is always a probability function that is "closer to the truth" in the sense of Brier score (and presumably any other proper scoring rule). I suspect that such results are a deal-breaker for the accuracy-first epistemologist, which I am not. But while I do not ascribe to that point of view, the result nevertheless raises the question of why one should give up the Bayesian ghost and "go non-additive" as Martin suggests. For Martin, the answer is clear as day: because of false confidence and the guarantee of validity that Martin–Liu belief functions provides.

Before I can jump on board with this, however, I still have to come to grips with the meaning of these belief functions $b_\bullet(\cdot)$ in the inferential setting, and furthermore with their role in helping me live up to the Statistician's Oath.

2.2.1 Truth or Belief?

Should the main component $b_{\bullet}(\cdot)$ of an inferential model report on the truth or the data analyst's beliefs? In Definition 1 it is said to report the data analyst's degree of belief in light of data $\bullet$. The validity condition then ties these beliefs to a conception of objective truth. I wonder if this can be related to Lewis's Principal Principle (PP), which plays a crucial role in reconciling subjective credences to objective chances in Bayesian epistemology, but I won't explore this connection here.

It seems obvious that one's beliefs about A should be as close to the 'truth' as possible, insofar as an assertion A can be regarded as 'true' or 'false'. But I'm afraid the desirability of this 'obvious' property is limited to philosophers and statisticians who are similarly insulated from the consequences that impact real-world statistical applications. Too often those who develop statistical methodology are far removed from the implementation, and thus consequences, of that methodology. Those applying these methods in real applications, however, have much more worldly concerns than an abstract notion of 'truth'.

First, in many statistical applications, the notion of 'truth' about A is pure fantasy. Consider a model $\mathcal{P}$ parameterized by a set Θ whose elements $\theta \in \Theta$ have no objective interpretation outside the context of the model $\mathcal{P}$. In other words, each θ determines a probability distribution $\mathbb{P}_\theta$, and those probabilities may be interpretable in terms of the predictions they make about an observable process, but beyond these predictions θ itself has no real world meaning. Statisticians obsess over doing inference about this made up parameter θ, constructing a world in which it is assumed that some such θ is 'true' and showing that in such a world the proposed method leads one to the truth with high probability. In such cases, I cannot divine the meaning of statistical inferences, such as in Martin's definition of IM that $b_y(A)$ is a degree of belief about the *truthfulness* of the assertion $\theta \in A$ because such an assertion is neither true nor false.

Second, and more down to earth, is that even in situations where the concept of 'truth' is well-understood, accuracy—having beliefs that are close to the truth—may not be a desirable criterion. Take the satellite collision application from Section 3.1. Any two satellites suspended in space either will or will not collide. There is a fact of the matter. Under the Brier score, the loss for having degree of belief p in a collision is $(1(A) - p)^2$, where $1(A)$ denotes the indicator of the event that the satellites collide. The loss is thus symmetric in the sense that assigning belief p to a collision that occurs suffers the same loss as assigning belief $1 - p$ to a collision that does not occur. However, and I think this is a key point in Martin's argument, these two outcomes are obviously not symmetric and shouldn't be treated as such.

Though I don't have experience in conjunction analysis or the logistics of satellites, I presume that a false belief of 99% that a collision will happen results in a relatively low-cost jittering of the satellite positions from a remote (earthly) location. A 1% false belief in a collision that does happen, however, leads to no action and, at best, damage to a satellite that remains in orbit or, at worst, the loss of an entire satellite. Clearly, the monetary costs of the two outcomes are on different scales.

From this perspective, it seems OK, even desirable, to skew one's beliefs toward incorrect assertions because of asymmetries associated with being wrong. Of course, every person knows this instinctively, which is why it is best practice to look both ways even before crossing a one way street. The cost of being taken out by a $< 0.01\%$ chance that a car is coming from the wrong direction far outweighs that of turning one's head to double check.

With this observation in mind, I find the appeal of the IM approach not to be avoiding "systematically misleading conclusions", as the Reid–Cox mantra suggests, but rather to live up to the Statistician's Oath. *Primum non nocere.* In fact, there are plenty of situations, satellite conjunction analysis being one, in which systematically *wrong* conclusions seem desirable, precisely because of the asymmetric costs and benefits involved. With this in mind, I'm not sure that the validity condition is sufficient to justify the use of IMs across the board. But I suspect that the condition might be generalized to account for the consequences associated with having certain beliefs, and to ensure that the model avoids dire consequences under all possible decisions drawn from the data. The refinement I have in mind would contain the validity condition as a (very) special case. It would also contain the Bayesian framework as a (very) special case.