Towards a Theory of Valid Inferential Models with Partial Prior Information

Ryan Martin[(✉)]

Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA
`rgmarti3@ncsu.edu`

Abstract. Inferential models (IMs) are used to quantify uncertainty in statistical inference problems, and *validity* is a crucial property that ensures the IM's reliability. Previous work has focused on validity in the special case where no prior information is available. Here I allow for prior information in the form of a non-trivial credal set, define a notion of validity and investigate its implications.

Keywords: Belief function · Coherence · Consonance · Generalized bayes · Possibility measure · Statistical inference

1 Introduction

In statistical inference, there are two dominant schools of thought: *Bayesian* and *frequentist*. The most significant difference between the two is that the former quantifies uncertainty about unknowns in a formal way, using the classical/ordinary/precise probability theory, while the latter does so in a less formal way, focusing on procedures—hypothesis tests, confidence sets, and other decision rules—that have appropriate control on their error rates. Numerous attempts, with different motivations, have been made to reconcile the two frameworks, including fiducial inference (e.g., Fisher 1935) and Dempster's extension (e.g., Dempster 1968), structural inference (Fraser 1968), generalized fiducial (Hannig et al. 2016), and confidence distributions (Schweder and Hjort 2016). Modern developments in this area are largely focused on the construction of data-dependent (precise) probability distributions from which procedures having frequentist error rate control properties (at least approximately) can be derived.

A different thread of work has focused on the development of data-dependent, imprecise probabilities that have a certain calibration or validity property designed to ensure that inferences drawn based on the magnitudes of the (lower and upper) probabilities would be reliable in a frequentist sense. Although ideas along similar lines appeared earlier in Balch (2012), to my knowledge, the first formal definition of *validity* and construction of an imprecise probability that achieves it was given in Martin and Liu (2013) and, later, in Martin and Liu

(2015); see, also, Martin (2019). Their construction of a valid *inferential model* (IM) makes use of random sets and, therefore, the imprecise probabilities take the form of (consonant) belief functions (Shafer 1976). These and other efforts to construct calibrated belief functions are surveyed in Denoeux and Li (2018).

In the spirit of de Finetti (1937), the focus in the imprecise probability literature is largely on the behavioral interpretation of the lower and upper probabilities; see Walley (1991) and Troffaes and de Cooman (2014). In particular, what minimal conditions on the mathematical structure of those lower and upper probabilities, treated as bounds on the prices an agent sets for gambles, are needed to protect him from sure loss? Since these *coherence* properties concern the internal reliability of the lower and upper probabilities, while the aforementioned validity property concerns a sort of external reliability, it makes sense to investigate the connections between the two.

After some brief background in Sect. 2, I give a definition of validity that is more general than those presented in the references above, and investigate its consequences in Sect. 3. In particular, I allow for available prior information in the form of a credal set—a collection of prior distributions—and present a definition of validity in such cases; previous work focus on the case where the credal set contains all possible priors. The motivation behind this extension is two-fold. First, the introduction of prior information brings the formulation closer to the subjective approach of de Finetti and Walley, where it's natural to consider behavioral implications, and I show in Proposition 2 that an agent adopting a pricing scheme based on lower and upper probabilities derived from a valid IM avoids sure loss. Second, in modern statistical problems involving high-dimensional unknowns, it's often believed that there's an underlying low-dimensional structure. These beliefs can be quantified using a set of prior distributions, so it's important to understand how the notion of validity might extend to such cases. I show that generalized Bayes is valid in the sense I defined. I also claim that a variation of Dempster's generalization of Bayesian inference would be valid too, but a precise statement and proof will be presented elsewhere.

However, it's important to emphasize that an IM being valid does not necessarily make it "good." For example, the IM could be inefficient in the sense that validity is achieved in a trivial way and the inferences drawn are not practically useful. In certain cases, especially those where little or no reliable prior information is available, there are other constructions—including one from Walley (2002) and one I refer to as "p-value + consonance"—that are more efficient without sacrificing efficiency. In cases where reliable prior information is available, efficiency can be gained by taking this into account. An open question is how to incorporate prior information so that both validity and this gain in efficiency is realized; see Cella and Martin (2019) for some first thoughts.

Finally, I present a notion of strong validity, which allows for a practically relevant uniformity over assertions, and I show that the approach advocated for in Martin (2019) and elsewhere achieves this stronger notion of validity and is also efficient, at least in the case of a vacuous prior.

For the sake of space, many details have been omitted. The full-length version (Martin 2021b), still in progress, contains proofs and more.

2   Problem Setup

Let Y denote observable data taking values in a sample space $\mathbb{Y}$; note that the sample space is general, so the data could be a vector, a matrix, etc. Next, consider a statistical model, $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$, a family of probability distributions on $\mathbb{Y}$ indexed by the parameter space Θ, which too is general. The goal is to quantify uncertainty about θ based on the observed data $Y = y$.

Prior information about θ might available in the form of a (closed and convex) credal set $\mathscr{Q}$ of prior distributions Q for θ. The "size" of $\mathscr{Q}$ controls the prior's precision, with $\mathscr{Q} = \{Q\}$ being the most precise and $\mathscr{Q} = \{$all probability distributions$\}$ being the least. These two extreme $\mathscr{Q}$'s are special: the former is classical Bayes while the latter matches the frequentist setup.

By uncertainty quantification, here I mean a data-dependent (precise or imprecise) probability distribution defined on a collection $\mathcal{A}$ of subsets of Θ. I will associate a subset $A \in \mathcal{A}$ with an assertion about the unknown, i.e., both A and "$\theta \in A$" will be called an *assertion*. Since the goal is to have something like a posterior distribution for θ, here I'll take $\mathcal{A}$ to be the Borel σ-algebra on Θ.

Following Martin (2019), define an *inferential model* (IM) as a mapping from data y, model $\mathscr{P}$, and prior information $\mathscr{Q}$ to a pair of lower and upper probabilities $(\underline{\Pi}_y, \overline{\Pi}_y)$ defined on $\mathcal{A}$. I'll interpret $\underline{\Pi}_y(A)$ and $\overline{\Pi}_y(A)$ as the y-dependent belief in and plausibility of the assertion A, respectively. It will be assumed throughout that $y \mapsto \overline{\Pi}_y(A)$ is Borel measurable for all $A \in \mathcal{A}$.

In the imprecise probability literature, it is common to give the lower and upper probabilities a behavioral interpretation. Imagine a situation where, after data $Y = y$ has been observed, the value of θ will be revealed and any gambles made on the truthfulness/falsity of assertions could be settled. Then the (subjective/personal) behavioral interpretation of my (data-dependent) lower and upper probabilities are

$$\underline{\Pi}_y(A) = \text{my maximum buying price for } 1(\theta \in A)$$
$$\overline{\Pi}_y(A) = \text{my minimum selling price for } 1(\theta \in A).$$

Here and in what follows, $1(E)$ denotes the indicator function of the event E. This behavioral interpretation, and one's clear desire to avoid being made a sure loser imposes certain constraints on the mathematical structure of the lower and upper probabilities. However, as I mentioned in Sect. 1, these mathematical constraints do not provide any assurance that the lower and upper probabilities are reliable in a statistical sense.

3   Statistical Properties

3.1   Validity

As discussed above, motivated by the behavioral interpretation, the imprecise probability literature mainly focuses on coherence. For data analysts on the

front lines, the ones crunching the numbers behind real-world decisions, this kind of internal rationality is important. From the perspective of a statistician who is developing methods for front-line data analysts to use off-the-shelf, there are other considerations. The only reason someone might use my method for their analysis is that they believe it's reliable, that it "works" in some specific sense. This goes beyond the internal rationality of coherence—lots of things that are coherent won't "work"—and this external rationality is what I call *validity*. A formal definition, more general than those in Martin and Liu (2013; 2015) and Martin (2019; 2021a), and its immediate consequences are below. These results extend the ideas developed by Cella and Martin (2020) in the context of prediction to cover the statistical inference problem.

First some additional notation. For the distribution P_θ of Y and a prior Q for θ, let P_Q denote the corresponding marginal distribution for Y and Q_y the corresponding conditional distribution of θ, given $Y = y$. Next, for a $Q \in \mathscr{Q}$, let M_Q denote the joint distribution of (Y, θ) under the corresponding Bayes model. Similarly, let $M_{\mathscr{Q}}$ denote the image of $\mathscr{Q}$ under $Q \mapsto M_Q$, and $\underline{M}_{\mathscr{Q}}$ and $\overline{M}_{\mathscr{Q}}$ as the lower and upper envelopes, respectively, corresponding to the assertion-wise infimum and supremum of M_Q over $Q \in \mathscr{Q}$. So, if E is any (appropriately measurable) joint event about (Y, θ), then the upper probability $\overline{M}_{\mathscr{Q}}(E)$ can be expressed more concretely as

$$\overline{M}_{\mathscr{Q}}(E) = \sup_{Q \in \mathscr{Q}} \iint 1\{(y, \theta) \in E\} \, P_\theta(dy) \, Q(d\theta)$$

$$= \sup_{Q \in \mathscr{Q}} \iint 1\{(y, \theta) \in E\} \, Q_y(d\theta) \, P_Q(dy).$$

Similarly, there is a corresponding lower probability, $\underline{M}_{\mathscr{Q}}$ that simply replaces the supremum above with an infimum, but this will not be used here.

Definition 1. *An IM $(\underline{\varPi}_Y, \overline{\varPi}_Y)$ is* valid, *relative to $(\mathscr{P}, \mathscr{Q})$, if either (and, hence, both) of the following equivalent conditions holds:*

$$\overline{M}_{\mathscr{Q}}\{\overline{\varPi}_Y(A) \leq \alpha, \, \theta \in A\} \leq \alpha, \quad \text{for all} (\alpha, A) \in [0, 1] \times \mathcal{A}, \qquad (1)$$

$$\overline{M}_{\mathscr{Q}}\{\underline{\varPi}_Y(A) > 1 - \alpha, \, \theta \notin A\} \leq \alpha, \quad \text{for all} (\alpha, A) \in [0, 1] \times \mathcal{A}. \qquad (2)$$

The equivalence of (1) and (2) follows from the duality $\overline{\varPi}_Y(A) = 1 - \underline{\varPi}_Y(A^c)$ and the "for all A" part of the conditions. The intuition behind this notion of validity is as follows. In applications, the data analyst will use the magnitudes of the IM's lower and upper probabilities to decide if the data support various assertions about θ. Of course, large values of $\underline{\varPi}_Y(A)$ support the truthfulness of A and small values $\overline{\varPi}_Y(A)$ support the truthfulness of A^c. So the events

$$\{(y, \theta) : \overline{\varPi}_y(A) \leq \alpha, \, \theta \in A\} \quad \text{and} \quad \{(y, \theta) : \underline{\varPi}_y(A) > 1 - \alpha, \, \theta \notin A\}$$

are situations when an erroneous conclusion may be made—or gamble may be lost—and the validity property controls the probability of these undesirable events, thus making the IM's uncertainty quantification reliable.

That this is a generalization of the valid inference framework presented in, say, Martin (2021a Definition 2), can be seen by considering the case where $\mathscr{Q}$ is the set of all probability distributions on Θ. In that case, validity in the sense of (1) reduces to

$$\sup_{\theta \in A} P_\theta\{\overline{\Pi}_Y(A) \leq \alpha\} \leq \alpha, \quad \text{for all}(\alpha, A) \in [0,1] \times \mathcal{A},$$

which is precisely the definition of validity in Martin (2021a).

A very basic requirement is that validity ought to imply that statistical procedures derived from the IM have certain error rate control guarantees. Proposition 1 below makes this precise.

Proposition 1. *Let $(\underline{\Pi}_Y, \overline{\Pi}_Y)$ be a valid IM in the sense of Definition 1. Then the following error rate control properties hold.*

1. A hypothesis testing rule that says reject *"$\theta \in A$"iff $\overline{\Pi}_Y(A) \leq \alpha$ satisfies*

$$\overline{M}_{\mathscr{Q}}\{\text{test rejects and}\theta \in A\} \leq \alpha.$$

2. The set $C_\alpha(y) = \bigcap\{A \in \mathcal{A} : \underline{\Pi}_y(A) > 1 - \alpha\}$ satisfies

$$\overline{M}_{\mathscr{Q}}\{C_\alpha(Y) \not\ni \theta\} \leq \alpha. \tag{3}$$

For some intuition about these results, consider two important (extreme) special cases corresponding to the traditional frequentist and Bayes approaches. For the frequentist case, where $\mathscr{Q}$ is all possible distributions, (3) immediately reduces to the familiar non-coverage probability bound, $\sup_\theta P_\theta\{C_\alpha(Y) \not\ni \theta\} \leq \alpha$, which is satisfied if C_α is a $100(1 - \alpha)\%$ confidence region in the traditional sense. Next, for the purely Bayes case, where $\mathscr{Q}$ is a singleton $\{Q\}$, $\overline{M}_{\mathscr{Q}}$ corresponds to a specific joint distribution of (Y, θ) and (3) is the condition automatically satisfied when C_α is the $100(1 - \alpha)\%$ posterior credible region.

Validity not only has implications for the operating characteristics of procedures derived from the IM, it also has behavioral implications. Proposition 2 below can be interpreted as saying that *validity implies no sure loss*. Avoiding sure loss is related to the aforementioned coherence properties (e.g., Walley 1991, Sect. 6.5.2), establishing a new perspective on validity compared to what had been discussed in previous works. This helps solidify the intuition that a procedure which is externally reliable shouldn't be internally irrational. The results below focus on the upper probability $\overline{\Pi}_y$; there are analogous properties expressed in terms of the corresponding lower probability $\underline{\Pi}_y$.

Proposition 2. *If an IM $(\underline{\Pi}_Y, \overline{\Pi}_Y)$ satisfies*

$$\sup_y \overline{\Pi}_y(A) < \underline{Q}(A) := \inf_{Q \in \mathscr{Q}} Q(A) \quad \text{for some } A, \tag{4}$$

then it's not valid, relative to $(\mathscr{P}, \mathscr{Q})$, in the sense of Definition 1.

A closer look at the validity property (1) reveals a relatively simple sufficient condition, namely, *dominance.* Indeed, by the iterated expectation formula,

$$\overline{M}_{\mathscr{Q}}\{\overline{\Pi}_Y(A) \leq \alpha, \, \theta \in A\} = \sup_{Q \in \mathscr{Q}} \int 1\{\overline{\Pi}_y(A) \leq \alpha\} \, Q_y(A) \, P_Q(dy), \quad (5)$$

so if $\overline{\Pi}_y(A) \geq Q_y(A)$ for all y, all $A \in \mathcal{A}$, and all $Q \in \mathscr{Q}$, then it follows immediately that (1) holds. But bounding an integral doesn't require uniform bounds on the integrand, it's enough for the above dominance to hold in an average sense. The following proposition makes this precise.

Proposition 3. *If the IM $(\underline{\Pi}_Y, \overline{\Pi}_Y)$ satisfies the following dominance property,*

$$\sup_{Q \in \mathscr{Q}} \int \frac{Q_y(A)}{\overline{\Pi}_y(A)} \, P_Q(dy) \leq 1, \quad \text{for all } A, \quad (6)$$

then it's valid, relative to $(\mathscr{P}, \mathscr{Q})$, in the sense of Definition 1.

An immediate consequence of Proposition 3 and the preceding discussion, if $\overline{\Pi}_y$ is the upper envelope in the generalized Bayes rule (e.g., Walley 1991, Sect. 6.4), that is, if

$$\overline{\Pi}_y(A) = \overline{Q}_y(A) := \sup_{Q \in \mathscr{Q}} Q(A \mid y), \quad (7)$$

then (6) holds and, therefore, so does validity in the sense of Definition 1. So the conservatism built in to the generalized Bayes rule, motivated by subjective coherence properties, is sufficient to achieve validity as well.

This also sheds light on what kinds of IM (likely) are not valid in the sense of Definition 1. For example, consider an approach like that described in Dempster (2008), where independent random sets/belief functions for θ—one based on prior information, the other based on data and statistical model—are combined, via Dempster's rule, to produce an IM $(\underline{\Pi}_Y, \overline{\Pi}_y)$. The probability intervals $[\underline{\Pi}_Y(A), \overline{\Pi}_Y(A)]$ obtained by Dempster's rule tend to be narrower than those corresponding to the generalized Bayes lower and upper envelopes (e.g., Kyburg 1987, Theorems A.3 and A.6). So, while I don't yet have a concrete counter-example at this time, the above sufficient condition generally doesn't hold, hence validity is questionable.

It's important to emphasize that dominance in the sense of (6) above is a sufficient but not necessary condition for validity. Indeed, there are other IM constructions besides the generalized Bayes lower/upper envelopes that are valid. One such construction is discussed below. Another is the combination of the prior-free IM for θ constructed in Martin and Liu (2015) with a prior belief function for θ via Dempster's rule; there's insufficient space to describe this here, so I'll present the result in a follow-up paper.

It's also worth emphasizing that validity, on its own, doesn't make the IM "good"—it may happen that (6) is achieved in a trivial way, which is not practically useful. For example, if the credal set $\mathscr{Q}$ is large, then the upper envelope

(7) in the generalized Bayes rule could be close to 1, for all/many A's, and then the inference would not be informative. So, beyond validity, it is necessary to consider the IM's *efficiency*.

3.2 Efficiency

As pointed out above, it's easy to see that the generalized Bayes solution is valid but perhaps in an inefficient, even trivial way. So, in a certain sense, the strong coherence properties satisfied by the generalized Bayes solution come at the cost of statistical efficiency. Since that formulation using lower and upper envelopes is only a sufficient condition for validity, there is an opportunity to find a more efficient solution, which is the focus of this subsection.

Towards finding a more efficient solution, let's consider a different strategy. If it can be shown that the IM's upper probability satisfies

$$\sup_{\theta} P_\theta \big\{ \overline{\Pi}_Y(\{\theta\}) \le \alpha \big\} \le \alpha, \quad \text{for all } \alpha \in [0,1], \tag{8}$$

then validity in the sense of Definition 1 holds by monotonicity:

$$\overline{\Pi}_Y(\{\theta\}) \le \alpha \implies \overline{\Pi}_Y(A) \le \alpha \text{ for all } A \ni \theta. \tag{9}$$

The condition in (8) is (roughly) what Walley (2002) calls the *fundamental frequentist principle*, or *FFP*; Walley's version says "$\alpha \in [0, \bar\alpha]$," for $\bar\alpha \le 1$. He then constructs an IM based on generalized Bayes applied to a special but broad credal set of the form $\mathscr{Q}_W = (1 - \varepsilon)\, Q_0 + \varepsilon\, \mathscr{Q}_{\text{all}}$, where Q_0 is a fixed prior distribution on Θ, $\mathscr{Q}_{\text{all}}$ is the set of all priors on Θ, and $\varepsilon \in (0, \frac{1}{2})$. Walley shows that the IM with upper probability $\overline{\Pi}_y = \overline{Q}_y$, with supremum over the special $\mathscr{Q}_W$, satisfies FFP which, for all practical purposes, implies validity in the sense of Definition 1 for all $\mathscr{Q}$, not just $\mathscr{Q}_W$. Most importantly, Walley's solution is far more efficient than, e.g., using generalized Bayes directly on $\mathscr{Q}_{\text{all}}$. However, as Walley notes, this solution is still inefficient in the sense that its plausibility intervals tend to be wider than classical confidence intervals.

A second option, more in line with the approach in Martin and Liu (2015) and Liu and Martin (2020), is as follows. Suppose one can find a function $\pi_y : \Theta \to [0, 1]$ with the property that the random variable $\pi_Y(\theta)$ satisfies the stochastic inequality in (8). This is precisely the property that typical p-values satisfy, so these functions are quite common. If that function also satisfies $\sup_\theta \pi_y(\theta) = 1$ for all y, then I can construct an IM whose upper probability is given by

$$\overline{\Pi}_y(A) = \sup_{\theta \in A} \pi_y(\theta), \quad A \in \mathcal{A}.$$

Under this construction, $\overline{\Pi}_y$ is a consonant plausibility function (Shafer 1976) or, equivalently, a possibility measure (Dubois and Prade 1988; Hose and Hanss 2021), and π_y is its corresponding *plausibility contour*. I'll refer to this below as the "p-value + consonance" IM construction. It's easy to show that, like Walley's

above, this IM is valid in the sense of Definition 1 for any $\mathscr{Q}$. However, this approach is generally more efficient than Walley's. For example, in a normal mean problem, Walley's plausibility interval has width of the order $(\log n)^{1/2} n^{-1/2}$, whereas the p-value + consonance intervals like in Martin and Liu (2015) have width of the order $n^{-1/2}$, just like classical confidence intervals.

A subtle point is that the meaning of "efficient" varies by the context. For example, when θ is relatively low-dimensional, the IM based on the construction in Martin and Liu (2015) is guaranteed to be valid and would generally be efficient. However, if θ is high-dimensional, then the same IM would tend to be inefficient. This is a sort of "curse of dimensionality"—increasing the dimension θ tends to inflate the plausibility function. More efficient solutions are possible when, as is typical in high-dimensional inference problems, there is an underlying low-dimensional structure. Combining this assumed low-dimensional structure/prior information with the data in an appropriate way would lead to a valid IM with improved efficiency compared to the no-prior IM. An open question is how to quantify and then incorporate that structural information so that both validity and efficiency are achieved? This will be answered elsewhere.

3.3   Strong Validity

While the validity condition in Definition 1 seems strong in the sense that it requires the inequalities (1–2) to hold for all assertions A, there is another sense in which it is too weak. In a gambling scenario, the agent will advertise his buying and selling prices based on his specified IM $(\underline{\Pi}_Y, \overline{\Pi}_Y)$, depending on data Y, and his opponents can decide what, if any, transactions they'd like to make. If the opponents also have access to data Y, then surely they will use that information to make a strategic choice of A in order to beat the agent. If the opponents can use data-dependent assertions, then it's not enough to consider the assertion-wise guarantees provided by Definition 1—some kind of uniformity in A is required. This scenario is not so far-fetched. Imagine a statistician who's developing a method for the applied data analyst to use. If the statistician can prove that his method satisfies (1–2), then his method is reliable for any fixed A. But what if the data analyst peeks at the data for guidance about relevant assertions? Without some uniformity, validity cannot be ensured in such cases. With this in mind, consider the following stronger notion of validity.

Definition 2. *An IM $(\underline{\Pi}_Y, \overline{\Pi}_Y)$ is* strongly valid, *relative to $(\mathscr{P}, \mathscr{Q})$, if*

$$\overline{M}_{\mathscr{Q}}\{\overline{\Pi}_Y(A) \le \alpha \text{ for some } A \ni \theta\} \le \alpha, \quad \text{for all } \alpha \in [0,1] \tag{10}$$

$$\overline{M}_{\mathscr{Q}}\{\underline{\Pi}_Y(A) > 1 - \alpha \text{ for some } A \not\ni \theta\} \le \alpha, \quad \text{for all } \alpha \in [0,1]. \tag{11}$$

Both Walley's and the p-value + consonance IM construction above achieve validity quite easily, arguably too easily. Perhaps this stronger notion of validity, with uniformity in A, is "just right." Indeed, it is not difficult to show that

$$\overline{\Pi}_y(A) \le \alpha \text{ for some } A \ni \theta \iff \overline{\Pi}_y(\{\theta\}) \le \alpha. \tag{12}$$

If the IM satisfies (8), which is akin to Walley's FFP, then strong validity follows.

Proposition 4. *If the IM's upper probability $\overline{\Pi}_Y$ satisfies* (8)*, then the strong validity property in Definition* 2 *holds for any $\mathscr{Q}$.*

The fact that the p-value + consonance construction presented here achieves strong validity and is generally more efficient than Walley's clever neighborhood model construction suggests that the former might be the "right" type of construction, and that IMs having this consonant structure are fundamental for statistical inference. I hope to present verification of these latter claims in follow-up work.

4   Conclusion

In this paper, I've investigated some more general version of the validity property first put forward in Martin and Liu (2015). An overarching goal of this and other ongoing work is to better understand the spectrum between the classical Bayesian setup with a single precise prior and the frequentist setup whose "prior" is completely imprecise/vacuous indexed by the level of imprecision, i.e., the size/complexity of $\mathscr{Q}$. Previous work had focused primarily on the latter frequentist setup and this paper gives a definition of validity that could be applied across a range of different precision levels in $\mathscr{Q}$.

The conclusion I draw from Proposition 4 above is that the p-value + consonance construction can be used to achieve (strong) validity for every $\mathscr{Q}$ and that, in a certain yet-to-be-formalized sense, is the "best" in the frequentist setup with a vacuous $\mathscr{Q}$; see, also, Martin (2021a). But this doesn't directly address the question of how to use genuine prior information in a non-extreme $\mathscr{Q}$ in a way that's both valid and efficient. Proposition 3 provides some minimal guidance, in particular, it says that generalized Bayes is a valid IM, but more investigation into its statistical efficiency is needed.

References

Balch, M.S.: Mathematical foundations for a theory of confidence structures. Internat. J. Approx. Reason. **53**(7), 1003–1019 (2012)

Cella, L., Martin, R.: Incorporating expert opinion in an inferential model while maintaining validity. In: De Bock, J., de Campos, C.P., de Cooman, G., Quaeghebeur, E., and Wheeler, G. (eds.) Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications. Proceedings of Machine Learning Research, vol. 103, pp. 68–77. PMLR, Thagaste, Ghent (2019)

Cella, L., Martin, R.: Validity, consonant plausibility measures, and conformal prediction. arXiv:2001.09225 (2020). https://doi.org/10.1016/j.ijar.2021.07.013

de Finetti, B.: La prévision: ses lois logiques, ses sources subjectives. Ann. Inst. H. Poincaré **7**(1), 1–68 (1937)

Dempster, A.P.: A generalization of Bayesian inference. (With discussion). J. Roy. Statist. Soc. Ser. B **30**, 205–247 (1968)

Dempster, A.P.: The Dempster-Shafer calculus for statisticians. Int. J. Approx. Reason. **48**(2), 365–377 (2008)

Denœux, T., Li, S.: Frequency-calibrated belief functions: review and new insights. Int. J. Approx. Reason. **92**, 232–254 (2018)

Dubois, D., Prade, H.: Possibility Theory. Plenum Press, New York (1988)

Fisher, R.A.: The fiducial argument in statistical inference. Ann. Eugen. **6**, 391–398 (1935)

Fraser, D.A.S.: The Structure of Inference. Wiley, New York (1968)

Hannig, J., Iyer, H., Lai, R.C.S., Lee, T.C.M.: Generalized fiducial inference: a review and new results. J. Am. Stat. Assoc. **111**(515), 1346–1361 (2016)

Hose, D., Hanss, M.: A universal approach to imprecise probabilities in possibility theory. Int. J. Approx. Reason. **133**, 133–158 (2021)

Kyburg, H.E., Jr.: Bayesian and non-Bayesian evidential updating. Artif. Intell. **31**(3), 271–293 (1987)

Liu, C., Martin, R.: Inferential models and possibility measures. arXiv:2008.06874 (2020)

Martin, R.: False confidence, non-additive beliefs, and valid statistical inference. Int. J. Approx. Reason. **113**, 39–73 (2019)

Martin, R.: An imprecise-probabilistic characterization of frequentist statistical inference (2021a). https://researchers.one/articles/21.01.00002

Martin, R.: Towards a theory of valid inferential models with partial prior information (2021b). https://researchers.one/articles/21.05.00001

Martin, R., Liu, C.: Inferential models: a framework for prior-free posterior probabilistic inference. J. Am. Stat. Assoc. **108**(501), 301–313 (2013)

Martin, R., Liu, C.: Inferential Models: Reasoning with Uncertainty. Monographs on Statistics and Applied Probability, vol. 147. CRC Press, Boca Raton (2015)

Schweder, T., Hjort, N.L.: Confidence, Likelihood, Probability. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 41. Cambridge University Press, New York (2016)

Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)

Troffaes, M.C.M., de Cooman, G.: Lower Previsions. Wiley Series in Probability and Statistics, Wiley, Chichester (2014)

Walley, P.: Statistical Reasoning with Imprecise Probabilities. Monographs on Statistics and Applied Probability, vol. 42. Chapman & Hall Ltd., London (1991)

Walley, P.: Reconciling frequentist properties with the likelihood principle. J. Stat. Plann. Inference **105**(1), 35–65 (2002)