Decision-theoretic properties of possibilistic inferential models

Ryan Martin[1]

[1]North Carolina State University, USA

Abstract

Inferential models (IMs) are data-dependent, imprecise-probabilistic structures designed to quantify uncertainty about unknowns. As the name suggests, the focus has been on uncertainty quantification for inference and on its reliability properties in that context. The present paper develops an IM framework for decision making, and investigates the decision-theoretic implications of the IM's reliability guarantees. I show that the IM's assessment of an action's quality, defined by a Choquet integral, will not be too optimistic compared to that of an oracle. This ensures that the IM tends not to favor actions that the oracle doesn't also favor, hence a IM is reliable for decision making too. In a certain special class of structured statistical models, further connections can be made between the IM's recommended actions and those recommended by Bayesian/fiducial frameworks, from which certain optimality conclusions can be drawn.

Keywords. Bayesian, confidence distribution, credal set, fiducial, group invariance, validity

1. Introduction

In some data analysis applications, the goal is to reliably infer features of a system under investigation. In others, the primary focus is on decision making. At least intuitively, quality decisions can be made without reliably inferring features of the system under investigation; but if one could reliably infer the relevant features of the underlying system, then good—perhaps even optimal—decisions ought to be within reach. The present paper's goal is to investigate this connection between reliable inference and formal decision-making.

Inferential model, or IMs for short, first introduced in Martin and Liu [46, 47], are data-dependent, (imprecise-) probabilistic structures designed to provide reliable uncertainty quantification about unknowns. The first version of the IM framework relied heavily on random sets and belief functions, but in this paper I'll focus on a more recent formulation that (largely) fits under the umbrella of possibility theory; accordingly, I refer to these as *possibilistic IMs* [e.g., 40, 43]. The specific reliability property that IMs—possibilistic or otherwise—satisfy is called *validity*; see [36, 37] and Section 2.1 below. Roughly, an IM is valid if its lower probabilities assigned to false hypotheses don't tend to be too large and, similarly, its upper probabilities assigned to true hypotheses don't tend to be too small. Among other things, validity also implies that IM-derived hypothesis tests and confidence regions exactly control frequentist error rates. It's precisely the validity property that distinguishes IMs from other brands of (imprecise) probabilistic statistical inference, including Bayesian, fiducial, and confidence distributions [8, 29, 68], Dempster–Shafer belief functions [14, 18, 56, 67], possibility measures [2, 20, 31], or more general kinds of imprecise probabilities [1, 65].

The IM developments described above are all focused on statistical inference; the formal decision problem remains to be addressed. To set the scene, a statistical decision problem comes equipped with a set $\mathbb{A}$ of possible actions, a collection $\mathbb{T}$ of possible states of the world, and a real-valued loss function $\ell_a(\theta)$ that represents the cost associated with taking action $a \in \mathbb{A}$ when the state of nature is $\theta \in \mathbb{T}$. If, in addition, a data-dependent probability distribution supported on $\mathbb{T}$ is available, then it's common to choose an action that minimizes the corresponding expected loss. Generalizations of this basic framework to cases where uncertainty is quantified via an imprecise probability have been developed, and I'll adopt the same approach here; that is, I'll rank actions according to a suitable upper expected loss, i.e., a Choquet integral (see Section 2.2). My main focus, however, is on the decision-theoretic implications of the IM's validity property. As stated in the first paragraph above, if the IM reliably solves the inference problem, then, in some sense, it should do the same for the decision problem. The paper's first key contribution, in Section 3, shows that the valid IM's upper expected loss tends to be not too small compared to that of a quasi-oracle who is in possession of partial information about the state of nature. This is relevant because, if the decision-maker is overly optimistic about the quality of an action, i.e., his assessment is more favorable than the quasi-oracle's, then there's a chance that he'll make a poor decision and incur a nontrivial loss. Theorem 3.1 shows, however, that validity implies the decision-maker's and quasi-oracle's assessments tend not to be inconsistent and, consequently, the valid IM is reliable for the decision problem too.

An important step in these developments is a characterization of the IM's credal set, i.e., the set of precise probabilities dominated by its upper probability. Existing results [e.g., 19, Prop. 4.1] imply that the IM's credal set consists of confidence distributions: for each $\alpha \in [0,1]$, assign probability at least $1-\alpha$ to the $100(1-\alpha)\%$ confidence sets. For models that have a special group transformation structure (Section 4), this characterization can be strengthened—the "maximal element" in the IM's credal set has a familiar form. Indeed, Fisher's fiducial argument is ideally suited for these models and the resulting fiducial distribution agrees with the Bayesian posterior distribution based on the the right invariant Haar prior; the common distribution returned by these two distinct arguments is a confidence distribution. Then [42] showed that the "maximal element" in the IM's credal set is exactly this Bayes/fiducial distribution. An important decision-theoretic consequence of this connection is Theorem 4.2, the paper's second main contribution, which says that, under certain conditions on the loss function, the IM's upper expected loss is minimized at the same action where the Bayes/fiducial distribution's expected loss is minimized. Combined with the main result in [60] about decision rules derived from the fiducial distribution, Theorem 4.2 implies the IM's recommended actions are generally high-quality and, in some cases, optimal. Therefore, while the IM's validity guarantees requires some degree of conservatism, this does not affect the quality of the IM's recommended actions.

2. Background

2.1. Inferential models. The first IM developments relied on random sets and belief functions. More recent developments in [40], building on [34, 35], define a possibilistic IM by applying a version of the probability-to-possibility transform to the model's relative likelihood. This paper will focus on the latter IM construction.

Consider a parametric model $\{\mathsf{P}_\theta : \theta \in \mathbb{T}\}$ consisting of probability distributions supported on a sample space $\mathbb{X}$, indexed by a parameter space $\mathbb{T}$. Suppose that the observable data X, taking values in $\mathbb{X}$, is a sample from the distribution P_Θ, where $\Theta \in \mathbb{T}$ is the unknown/uncertain "true value." The model and observed data $X = x$ together determine a likelihood function $\theta \mapsto p_\theta(x)$ and a corresponding relative likelihood

$$R(x,\theta) = \frac{p_\theta(x)}{\sup_\vartheta p_\vartheta(x)}.$$

I will implicitly assume here that the denominator is finite for almost all x. As is typical in the literature, I will also assume that prior information about Θ is vacuous.

The relative likelihood defines a possibility contour, i.e., a non-negative function such that $\sup_\theta R(x,\theta) = 1$ for almost all x. This contour determines a possibility measure that can be used for data-driven uncertainty quantification about Θ, which has been extensively studied in [e.g., 15, 16, 57, 67]. This likelihood-driven possibility has a number of desirable properties. What it lacks, however, is a justification for why the "possibilities" assigned to hypotheses about Θ carry sufficient weight to (in)form the data analyst's beliefs. With vacuous prior information, there's no Bayesian justification behind these possibility assignments, so justification can only come from a validity-like frequentist reliability guarantee. But the likelihood-based possibility assignment falls short of this goal. So, while the relative likelihood provides a useful, data-driven parameter ranking, I'd argue that this is insufficient for reliable statistical inference.

Fortunately, it's straightforward to achieve the desired goal by "validifying" [39] the relative likelihood. This amounts to applying a version of the probability-to-possibility transform [e.g., 21, 30], and the result is the possibilistic IM's contour function:

$$\pi_x(\theta) = \mathsf{P}_\theta\{R(X,\theta) \le R(x,\theta)\}, \quad \theta \in \mathbb{T}. \tag{1}$$

When a suitable ancillary statistic (one whose sampling distribution doesn't depend on unknown model parameters) can be identified, my recommendation in [40]—akin to recommendations by Fisher [23], Berger [6], and others—is to reduce the complexity/dimension by conditioning on the observed value of the ancillary statistic in the above P_θ-probability calculation; see Section 4 below. Then the corresponding possibility measure, or upper probability, is defined via optimization as

$$\overline{\Pi}_x(H) = \sup_{\theta \in H} \pi_x(\theta), \quad H \subseteq \mathbb{T}. \tag{2}$$

It won't be needed in the present paper, but the corresponding necessity measure, or lower probability is $\underline{\Pi}_x(H) = 1 - \overline{\Pi}_x(H^c)$. As in Section 1, an essential feature of this IM construction is its *validity* property:

$$\sup_{\Theta \in \mathbb{T}} \mathsf{P}_\Theta\big\{\pi_X(\Theta) \le \alpha\big\} \le \alpha, \quad \alpha \in [0,1]. \tag{3}$$

(Remember that Θ is the true parameter value; so what I mean by "$\sup_\Theta \mathsf{P}_\Theta$" is probability bound holds for any true value.) Property (3) has a number of important consequences. First, it immediately implies that

$$C_\alpha(x) = \{\theta \in \mathbb{T} : \pi_x(\theta) \ge \alpha\}, \quad \alpha \in [0,1] \tag{4}$$

is a $100(1-\alpha)\%$ frequentist confidence set, i.e., $\sup_{\Theta \in \mathbb{T}} \mathsf{P}_\Theta\big\{C_\alpha(X) \not\ni \Theta\big\} \le \alpha$. Second, from (2) and (3),

$$\sup_{\Theta \in H} \mathsf{P}_\Theta\big\{\overline{\Pi}_X(H) \le \alpha\big\} \le \alpha, \quad \alpha \in [0,1], H \subseteq \mathbb{T}. \tag{5}$$

In words, a valid IM assigns possibility $\le \alpha$ to true hypotheses at rate $\le \alpha$ as a function of data X. This gives the IM its "inferential weight"—(5) implies that $\overline{\Pi}_x(H)$ is not expected to be small when H is true, so one is inclined to doubt the truthfulness of a hypothesis H if $\overline{\Pi}_x(H)$ is small. Third, the above property ensures that the possibilistic IM is safe from false confidence [3, 36, 44], unlike default-prior Bayes and fiducial solutions. Further properties of possibilistic IMs' are available, including de Finetti-style no-sure-loss, uniform validity, and asymptotic efficiency [12, 40, 41, 48].

The IM output is a genuine/coherent imprecise probability and, consequently, there is a set of precise probability distributions that are compatible with it. This set is called the *credal set* associated with the possibilistic IM's output $\overline{\Pi}_x$, and is expressed mathematically as

$$\mathscr{C}(\overline{\Pi}_x) = \{\mathsf{Q}_x \in \text{probs}(\mathbb{T}) : \mathsf{Q}_x(\cdot) \leq \overline{\Pi}_x(\cdot)\}, \qquad (6)$$

where probs($\mathbb{T}$) is the set of probabilities supported on (the Borel σ-algebra of subsets of) $\mathbb{T}$ and "$\mathsf{Q}_x(\cdot) \leq \overline{\Pi}_x(\cdot)$" means the inequality holds for all events on $\mathbb{T}$. Of course, the members of $\mathscr{C}(\overline{\Pi}_x)$ depend on x because $\overline{\Pi}_x$ does, but these may not correspond to Bayesian posterior distributions under any prior. Fortunately, an interpretation can be given to the members of $\mathscr{C}(\overline{\Pi}_x)$ thanks to a well-known characterization [e.g., 19, Prop. 2.1]:

$$\mathsf{Q}_x \in \mathscr{C}(\overline{\Pi}_x) \iff \mathsf{Q}_x\{C_\alpha(x)\} \geq 1 - \alpha, \text{ all } \alpha \in [0, 1],$$

where $C_\alpha(x)$ is as defined in (4) with π_x the contour corresponding to $\overline{\Pi}_x$. Since $C_\alpha(x)$ is a $100(1-\alpha)\%$ confidence set, and the elements Q_x in the credal set assign at least probability $1-\alpha$ to $C_\alpha(x)$, there's good reason to call these elements *confidence distributions*. This definition of confidence distributions agrees with that given in [59] and generalizes those commonly found in the statistics literature [13, 54, 68]. This credal set characterization and its interpretation as a collection of confidence distributions will be relevant below.

While the IM construction is conceptually simple and its properties are strong, computation can be a challenge. My go-to strategy has been

$$\pi_x(\theta) \approx \frac{1}{M} \sum_{m=1}^{M} 1\{R(X_{m,\theta}, \theta) \leq R(x, \theta)\}, \quad \theta \in \mathbb{T},$$

where $X_{m,\theta}$ are independent copies of the data X, drawn from P_θ, for $m = 1, \ldots, M$. This is feasible at a few different θ values, but, e.g., evaluation over a fine grid covering the relevant portion of the parameter space $\mathbb{T}$ can be expensive. Recent developments [45] have revealed new opportunities for much more efficient computation.

2.2. Decision theory. Decision theory aims to describe the "optimal" behavior for an agent faced with a decision problem. This requires comparing the agent's strategies and ranking them by preference. Early efforts [9] went directly to considerations of loss (or negative utility), ranking strategies according to their expected loss and, hence, defining the optimal strategy as one that minimizes expected loss. Other authors took a different route, by considering general preference orders on strategies, but both reached effectively the same destination. For example, the celebrated theorem of Neumann and Morgenstern [50] says that if the preference order satisfies certain rationality axioms, then there exists a loss function such that the ranking of strategies according to preferences is equivalent to the ranking by expected loss. Similar conclusions have been reached by others [52], hence, the *minimize-expected-loss* principle.

For the statistical decision problem, let $(\theta, a) \mapsto \ell_a(\theta)$ denote a loss function that measures the loss incurred by taking an action $a \in \mathbb{A}$ when the "state of nature" is θ. Common examples include squared-error loss, with $\ell_a(\theta) = \|a - \theta\|^2$, and 0–1 loss, with

$$\ell_a(\theta) = 1(\theta \in H, a = 1) + 1(\theta \notin H, a = 0),$$

where $H \subset \Theta$ is a (null) hypothesis, and $a = 1$ and $a = 0$ denote "reject" and "do not reject" H, respectively. I'll assume throughout that the loss is non-negative, which is not unreasonable. While negative losses might make sense in some contexts, these gains would typically be bounded and so the loss could be made non-negative by adding a constant. This additive constant won't affect judgments of the relative quality of actions.

Of course, an oracle who knows the true Θ could easily compare possible actions according to their corresponding loss values, $\ell_a(\Theta)$. I'll refer to $a \mapsto \ell_a(\Theta)$ as the *oracle's assessment* of action a. Naturally, the oracle would take the action $a^\star = \arg\min_a \ell_a(\Theta)$ and incur minimal loss. None of us have oracle powers, so $\ell_a(\Theta)$ is unknown and a different approach is required. When uncertainty about Θ, given $X = x$, is quantified via a probability Q_x, like a Bayesian posterior distribution, a typical strategy is to rank the candidate actions according to their expected loss, $a \mapsto \mathsf{Q}_x \ell_a$. Then the obvious analogue to the known-Θ case mentioned above is to minimize the expected loss, $\hat{a}(x) = \arg\min_a \mathsf{Q}_x \ell_a$.

Other authors have argued that requiring a single precise probability to quantify uncertainty puts too much of a burden on the data analyst, and that a decision-theoretic framework based on more general imprecise probabilities is better suited for practical applications. Excellent reviews of decision theory from an imprecise probability perspective can be found in [32] and [17]. Roughly, given a loss function like above, if uncertainty is quantified via an imprecise probability, then it's only natural to extend the minimize-expected-loss principle by replacing the expected loss, $\mathsf{Q}_x \ell_a$, with an appropriate generalization, and then optimizing to find the best action [e.g., 25, 26]. It was shown in [27] that an appropriate generalization of the expected loss is obtained via the

Choquet integral; see Appendix C of [63] for full details, Section 4.1 of [17] for a summary, and Section 3 below. Under an imprecise probability framework, there are two versions of "expected loss," an upper and a lower, and I'll define the optimal action to be that which minimizes the upper expected loss. This strategy is commonly referred to as *minimax*, and there is extensive work about this in the literature on imprecise probability [32] and in robust Bayesian inference [4, 5, 64].

3. IMs and decision-making

3.1. Setup. I'll start by developing the decision-theoretic framework for IMs. When the upper probability $\overline{\Pi}_x$ is interpreted as an upper envelope on a collection of ordinary probabilities, then, naturally, its extension to an upper prevision/expectation is

$$\overline{\Pi}_x f = \sup\{\mathsf{Q}_x f : \mathsf{Q}_x \in \mathscr{C}(\overline{\Pi}_x)\}, \tag{7}$$

where $\mathscr{C}(\overline{\Pi}_x)$ is the credal set (6) and f is a suitable real-valued function defined on $\mathbb{T}$. In light of the discussion in Section 2.1, the upper envelope has some further intuition, that is, $\overline{\Pi}_x f$ is largest of the ordinary expected values, $\mathsf{Q}_x f$, corresponding to confidence distributions Q_x compatible with $\overline{\Pi}_x$. There is an associated lower prevision, $\underline{\Pi}_x f$, but this won't be needed.

The above extension of an upper probability to an upper prevision can be equivalently described as a *Choquet integral*. The focus here is on IMs whose upper probability $\overline{\Pi}_x$ is a possibility measure. For these, Propositions 7.14 and 15.42 in [63], in increasing generality, establish that the Choquet integral of a non-negative function $f : \mathbb{T} \to [0, \infty)$ with respect to the possibility measure $\overline{\Pi}_x$, if it exists, is given by

$$\overline{\Pi}_x f = \int_0^1 \Big\{ \sup_{\theta : \pi_x(\theta) > s} f(\theta) \Big\} \, ds, \tag{8}$$

where π_x is the possibility contour corresponding to $\overline{\Pi}_x$. Existence of the Choquet integral requires that the function f be "previsible" [63, Def. 15.6], which I'll silently assume about the loss functions below. Importantly, the Choquet integral determines the upper prevision, hence the right-hand sides of (7) and (8) are equal.

Since the loss function is assumed to be non-negative, either of the above two equivalent formulas can be applied directly to define an upper risk/expected loss,

$$a \mapsto \overline{\Pi}_x \ell_a. \tag{9}$$

Assuming $\overline{\Pi}_x \ell_a$ is finite at least for some actions a, this can be used to assess the quality of different actions, relative to the given loss function and the IM's data-dependent possibility measure. In particular, one can select a x-dependent "optimal" action as

$$\hat{a}(x) = \arg\min_a \overline{\Pi}_x \ell_a, \tag{10}$$

which has a *minimax* connotation, being the action that minimizes the upper expected loss. The rationale behind the use of upper instead of lower expected loss is conservatism, i.e., if an action has small upper expected loss, then, by (7), the corresponding expected loss with respect to any compatible $\mathsf{Q}_x \in \mathscr{C}(\overline{\Pi}_x)$ is also small. As mentioned in Section 2.2 and in more detail in [17], the ranking of actions based on magnitudes of the upper expected loss is consistent with certain rationality axioms. Finally, while the minimax strategy I'm taking is reasonable, it's not the only such strategy; alternatives include *maximax* and *e-admissibility*, as described in [e.g., 32, 55, 62], but I will not consider these strategies here.

3.2. Properties. Recall that, roughly, validity ensures that the IM's user won't make systematically misleading inferences. In the present context, a poor choice of action may result if the IM's assessment of a, based on $\overline{\Pi}_x \ell_a$ in (9), were very different from the oracle's assessment $\ell_a(\Theta)$ where, again, Θ is the unknown true parameter. That is, a user hopes to avoid cases where, for some action a, their assessment $\overline{\Pi}_x \ell_a$ in (9) of the loss associated with action a is much more optimistic than the oracle's, i.e., $\overline{\Pi}_x \ell_a \ll \ell_a(\Theta)$. Again, such cases are undesirable because a situation where the user's assessment of an action is much more optimistic than the oracle's creates an unfortunate opportunity for the user to choose a poor action and suffer a non-trivial loss.

A first question is if the IM and oracle assessments are even comparable. Towards a quick, non-rigorous, affirmative answer to this question, suppose that the data $X = X^n$ consists of n iid observations from P_Θ. Then the asymptotic properties proved in Martin and Williams [48] imply that π_{X^n} will, with P_Θ-probability converging to 1, collapse to $\theta \mapsto 1(\theta = \Theta)$ as $n \to \infty$. So, at least for actions a such that $\theta \mapsto \ell_a(\theta)$ is continuous in a neighborhood of Θ, it follows that $\overline{\Pi}_{X^n} \ell_a \to \ell_a(\Theta)$.

Given that the two assessments are generally comparable, it makes sense to be more precise about what it means for the IM's assessment, $\overline{\Pi}_x \ell_a$, of action a to be "much more optimistic" than the oracle's assessment, $\ell_a(\Theta)$. Define the data-dependent local maximum loss,

$$\begin{aligned} L_a(x, \Theta) &= \sup\{\ell_a(\theta) : \theta \in C_{\pi_x(\Theta)}(x)\} \\ &= \sup\{\ell_a(\theta) : \pi_x(\theta) > \pi_x(\Theta)\}, \end{aligned} \tag{11}$$

where the second equality follows by the definition (4) of the plausibility region $C_\alpha(x)$. I'll refer to $L_a(x, \Theta)$ as the *quasi-oracle's assessment* of a, based on data $X = x$. This corresponds to an oracle who doesn't know the exact value of Θ but knows which values θ are at least as π_x-plausible as Θ relative to data $X = x$. Then by "much more optimistic" I mean $\overline{\Pi}_x \ell_a$ being less than a small multiple of $L_a(x, \Theta)$, i.e.,

$$\overline{\Pi}_x \ell_a \le \alpha \, L_a(x, \Theta), \quad \alpha \in (0, 1).$$

It's easy to see from (11) that $L_a(x,\Theta) \geq \ell_a(\Theta)$, for all x, so the quasi-oracle's assessment is more conservative than the oracle's and possibly within reach. Therefore, if my IM's assessment, $\overline{\Pi}_x \ell_a$, of the loss associated with action a tends not to be much more optimistic than the quasi-oracle's assessment, $L_a(x,\Theta)$, then it also won't tend to be much more optimistic than the oracle's assessment, $\ell_a(\Theta)$. And not tending to be much more optimistic than the oracle provides some assurance that poor actions won't be favored by the IM.

Below is (to my knowledge) the first result showing the implications of validity in the context of decision theory. It says that the valid possibilistic IM's assessment $\overline{\Pi}_X \ell_a$ tends to be not considerably smaller than $L_a(X,\Theta)$ for all actions $a \in \mathbb{A}$ and for any true $\Theta \in \mathbb{T}$.

Theorem 3.1. *Let $\ell_a : \mathbb{T} \to [0,\infty)$ be a non-negative loss function for each a in the action space $\mathbb{A}$. For an upper probability $\overline{\Pi}_x$, define the minimum ratio*

$$\mathcal{R}(x,\Theta) = \inf_{a\in\mathbb{A}} \frac{\overline{\Pi}_x \ell_a}{L_a(x,\Theta)},$$

where $L_a(x,\Theta)$ is as in (11). If the IM with upper probability $\overline{\Pi}_X$ is valid in the sense of (3), then

$$\sup_\Theta \mathsf{P}_\Theta\{\mathcal{R}(X,\Theta) \leq \alpha\} \leq \alpha, \quad \textit{for all } \alpha \in [0,1],$$

where the supremum implies that the probability bound holds for any true value Θ.

Proof. Define the function

$$h_{x,a}(s) = \sup_{\theta:\pi_x(\theta)>s} \ell_a(\theta), \quad s \in [0,1], \tag{12}$$

so that the Choquet integral $\overline{\Pi}_x \ell_a$ is just a Riemann integral of $h_{x,a}$. It's clear that $s \mapsto h_{x,a}(s)$ is decreasing, which implies

$$\begin{aligned}\overline{\Pi}_x \ell_a &= \int_0^1 h_{x,a}(s)\,ds \\ &= \Bigl(\int_0^{\pi_x(\Theta)} + \int_{\pi_x(\Theta)}^1\Bigr) h_{x,a}(s)\,ds \\ &\geq \pi_x(\Theta)\, h_{x,a}\bigl(\pi_x(\Theta)\bigr).\end{aligned}$$

This is a Markov inequality for the Choquet integral [66]. Since $L_a(x,\Theta) = h_{x,a}(\pi_x(\Theta))$, it follows that

$$\begin{aligned}\mathcal{R}(X,\Theta) \leq \alpha &\iff \overline{\Pi}_X \ell_a \leq \alpha\, L_a(X,\Theta) \quad \text{some } a \\ &\implies \pi_X(\Theta)\, L_a(X,\Theta) \leq \alpha\, L_a(X,\theta) \quad \text{some } a \\ &\iff \pi_X(\Theta) \leq \alpha.\end{aligned}$$

Validity implies the latter event has P_Θ-probability $\leq \alpha$, uniformly in Θ, which proves the claim. ☐

I presented the above theorem with the infimum over actions $a \in \mathbb{A}$ because making the uniformity explicit makes the strength of the result clear. But, admittedly, the uniformity makes the result more difficult to interpret. One relevant consequence is presented in the following corollary, where instead of the infimum of risk assessments over all actions, I plug in a specific action, namely, the upper expected loss minimizer $\hat{a}(X)$.

Corollary 3.1. *Under the setup of Theorem 3.1, if $\hat{a}(x) = \arg\min_a \overline{\Pi}_x \ell_a$ is the IM's optimal action, then*

$$\sup_\Theta \mathsf{P}_\Theta\{\overline{\Pi}_X \ell_{\hat{a}(X)} \leq \alpha\, L_{\hat{a}(X)}(X,\Theta)\} \leq \alpha, \quad \alpha \in [0,1].$$

Proof. The infimum ratio over actions $a \in \mathbb{A}$ can be no larger than the ratio at a specific (data-dependent) action $\hat{a}(x)$. Then the claim follows from Theorem 3.1. ☐

The "quasi-oracle risk assessment" complicates interpretation but, fortunately, it's easy to remove this complication by replacing the quasi-oracle assessment $L_a(X,\Theta)$ with the simpler oracle assessment $\ell_a(\Theta)$.

Corollary 3.2. *Under the setup of Theorem 3.1,*

$$\sup_\Theta \mathsf{P}_\Theta\Bigl\{\inf_{a\in\mathbb{A}} \frac{\overline{\Pi}_X \ell_a}{\ell_a(\Theta)} \leq \alpha\Bigr\} \leq \alpha, \quad \alpha \in [0,1].$$

Furthermore, just as in Corollary 3.1, if $\hat{a}(X)$ is the upper risk minimizer, then the IM's risk assessment satisfies

$$\sup_\Theta \mathsf{P}_\Theta\{\overline{\Pi}_X \ell_{\hat{a}(X)} \leq \alpha\, \ell_{\hat{a}(X)}(\Theta)\} \leq \alpha, \quad \alpha \in [0,1].$$

Proof. This claim follows by Theorem 3.1 and the simple fact that $\ell_a(\theta) \leq L_a(x,\theta)$ for all (x,θ,a). ☐

For a quick recap, recall that it would be undesirable if the IM's assessment of the quality of a were much more optimistic than the oracle's or the quasi-oracle's—it would put the decision-maker at risk of suffering non-trivial loss. This is especially true for the "best" action, $a = \hat{a}(x)$, the one that the decision-maker is likely to take. By Theorem 3.1, the IM's validity implies that such undesirable cases are rare events with respect to the distribution of X. Therefore, validity provides some assurance that the IM's data-driven assessment of action a is not inconsistent with that of the oracle or quasi-oracle, hence the IM helps the decision-maker mitigate risk.

3.3. Illustration. Consider a simple location model, on $\mathbb{X} = \mathbb{R}$, where the density p_θ corresponding to the distribution P_θ is of the form $p_\theta(x) = p(x-\theta)$, for some density p symmetric around 0, e.g., Gaussian, Student-t, etc. I'll assume that $n = 1$, but neither this nor symmetry of p are necessary; this just makes the calculations easy to do by hand and helps reveal some interesting structure that will be investigated in the next section.

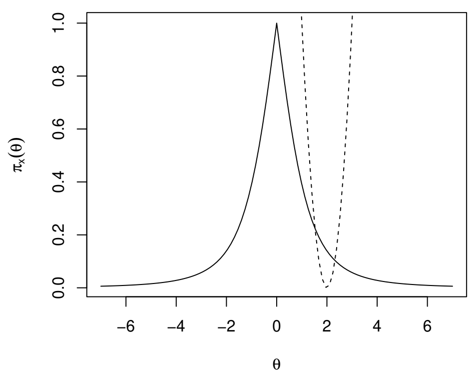

Figure 1. *Plot of $\pi_x(\theta)$, for $x = 0$, when p is a Student-t density with 3 degrees of freedom; dashed line shows $\ell_a(\theta)$, when $a = 2$.*

Suppose that $X = x$ is the observed data. From symmetry, it's easy to see that the possibilistic IM contour for Θ, at x, as defined in (1), reduces to

$$\begin{aligned}\pi_x(\theta) &= \mathsf{P}_\theta\{p(X-\theta) \le p(x-\theta)\} \\ &= \mathsf{P}_\theta\{|X-\theta| \ge |x-\theta|\} \\ &= 2\{1 - P(|x-\theta|)\}, \qquad \theta \in \mathbb{R},\end{aligned}$$

where P is the distribution function corresponding to the density p. Then it's easy to see that the $100(1-\alpha)\%$ confidence interval (4) derived from the IM is

$$C_\alpha(x) = \big[x - P^{-1}(1-\tfrac{\alpha}{2}),\, x + P^{-1}(1-\tfrac{\alpha}{2})\big].$$

To define a decision problem, consider "point estimation" where the action space $\mathbb{A}$ is the parameter space, and the goal is to select an action $\hat{a}(x)$, depending on data x, that's best in the sense of minimizing $a \mapsto \overline{\Pi}_x \ell_a$ for a suitable loss function ℓ_a. As is customary, I'll use the squared error loss, i.e., $\ell_a(\theta) = (a-\theta)^2$. It'll be clear in what follows that the upper risk $\overline{\Pi}_x \ell_a$ is well-defined only if the density p has sufficiently thin tails that its variance $v(p)$ is finite, so I'll assume this throughout.

Figure 1 shows both the contour function $\pi_x(\theta)$ and the loss function $\ell_a(\theta)$ for specific values of x and a, with p a Student-t density. Note, in particular, that π_x is directionally convex and, hence, on any level set of π_x, the loss function attains its maximum at the boundary. Using this observation, and applying the Choquet integral formula in (8), I get

$$\begin{aligned}\overline{\Pi}_x \ell_a &= \int_0^1 \sup_{\theta : 2\{1-P(|x-\theta|)\} > s} (a-\theta)^2 \, ds \\ &= \int_0^1 \max_{+,-} \{x - a \pm P^{-1}(1-\tfrac{s}{2})\}^2 \, ds \\ &= \int_0^1 \{|x-a| + P^{-1}(1-\tfrac{s}{2})\}^2 \, ds \\ &= \int_{-\infty}^{\infty} \big(|x-a| + |z|\big)^2 \, p(z)\, dz \\ &= (x-a)^2 + 2|x-a|\, m(p) + v(p),\end{aligned}$$

where $m(p) = \int |z|\, p(z)\, dz$. For comparison, there is a standard fiducial/confidence distribution for the location problem, which is also the Bayesian posterior distribution under the default, flat prior for Θ. That posterior, denoted by $\mathsf{Q}_x^\star$, has density function $q_x^\star(\theta) = p(\theta - x)$, and it's easy to check that the corresponding fiducial expected loss is $\mathsf{Q}_x^\star \ell_a = (x-a)^2 + v(p)$. Note that $\mathsf{Q}_x^\star \ell_a \le \overline{\Pi}_x \ell_a$ for all a, with equality if and only if a equals x. Indeed, the two expected loss functions are equal at their common minimizer $\hat{a}(x) = x$. This minimizer is Pitman's optimal equivariant estimator, also the Bayes estimator under a flat prior. As I show in Section 4, it's not a coincidence that the IM's suggested action based on the minimum upper risk principle aligns with that of other methods in this location model.

It's also of interest to consider the IM's risk assessment compared to the oracle's. The result in Theorem 3.1 is quite simple and general, which means that it can be conservative in certain instantiations. It so happens that a stronger result can be shown in scalar parameter cases like the one under consideration here. I won't present the full details here; but see Theorem 3 and the related discussion in [38]. Under some mild conditions satisfied in this example, the conclusion in Theorem 4.2 holds with the quasi-oracle's assessment $L_a(X, \Theta)$ replaced by

$$\widetilde{L}_a(X,\Theta) := \sup\{\ell_a(\theta) : \pi_X(\theta) \ge \tfrac{1}{2}\pi_X(\Theta)\}.$$

For illustration, define $\widetilde{\mathcal{R}}(X,\Theta)$ as the ratio corresponding to $\mathcal{R}(X,\Theta)$ but with $\widetilde{L}$ in the denominator. Figure 2 shows the distribution function of $\widetilde{\mathcal{R}}(X,\Theta)$, i.e.,

$$r \mapsto \mathsf{P}_\Theta\{\widetilde{R}(X,\Theta) \le r\}, \quad r \in [0,1], \tag{13}$$

and the fact that this curve is below the diagonal confirms the above claim. For comparison, Figure 2 also shows the distribution function of the random variable

$$\inf_{a \in \mathbb{A}} \frac{\mathsf{Q}_X^\star \ell_a}{\widetilde{L}_a(X,\Theta)},$$

where $\mathsf{Q}_X^\star$ is the fiducial/Bayes posterior distribution described above. As is clear from the plot, the above claim does not hold for the fiducial/Bayes assessment of the risk. That is, $\mathsf{Q}_X^\star \ell_a$ tends to be a bit too optimistic compared to $\widetilde{L}_a(X,\Theta)$. Whether this difference in fiducial/Bayes and IM performance has practically relevant consequences remains to be seen.

3.4. Application. Random effect models are ubiquitous in applications across business, engineering, and science. The simplest such model is commonly written as

$$X_{ij} = \theta_0 + T_i + E_{ij}, \quad i = 1,\dots,m, \quad j = 1,\dots,n,$$

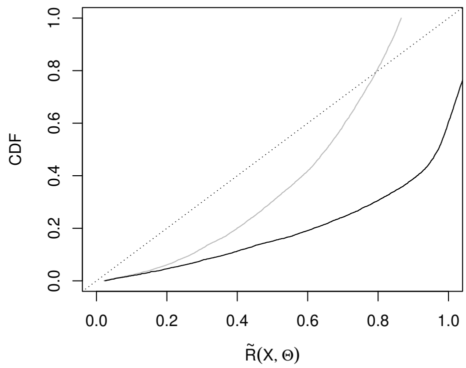

Figure 2. *The black line shows the distribution function (13) of $\widetilde{\mathcal{R}}(X, \Theta)$ for the symmetric location model IM. The gray line shows the corresponding distribution function for the fiducial/Bayes version of $\widetilde{\mathcal{R}}(X, \Theta)$. The diagonal line corresponds to the* $\mathsf{Unif}(0,1)$ *distribution function.*

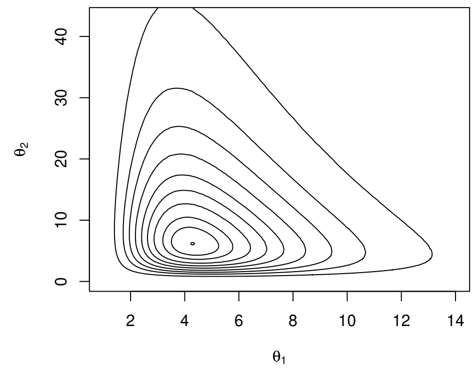

Figure 3. *Plots of the possibilist IM contour for the variance components (θ_1, θ_2).*

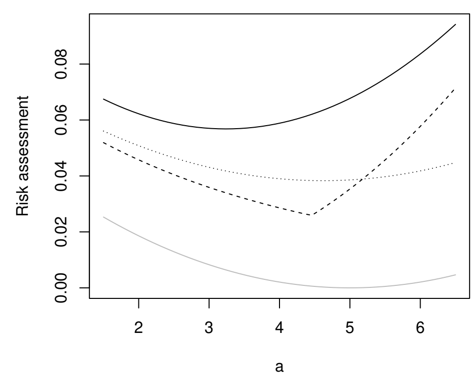

Figure 4. *Plots of the oracle (gray), likelihood (dashed), Bayes (dotted), and IM (solid) risk assessments based on loss function (14).*

where $X = (X_{ij})$ is the observable data and T_i and E_{ij} are mutually independent random effects, with

$$E_{ij} \overset{\text{iid}}{\sim} \mathsf{N}(0, \theta_1) \quad \text{and} \quad T_i \overset{\text{iid}}{\sim} \mathsf{N}(0, \theta_2).$$

Then the three-dimensional parameter that indexes this model is $\theta = (\theta_0, \theta_1, \theta_2)$, where θ_0 is the overall mean and (θ_1, θ_2) are variance components associated with error/replication and treatment, respectively. The overall mean is not of primary interest, and it can be easily marginalized out, so I'll assume that this has been done throughout. The focus will be on the variance components, in particular, that for the treatment effect.

The work that follows is all based on a simulated data set with $m = 5$, $n = 3$, and true parameters $(\Theta_0, \Theta_1, \Theta_2) = (0, 7, 5)$; the methods that follow depend on the data on through the statistics $S_1 = \sum_i \sum_j (X_{ij} - \bar{X}_i)^2 = 42.94$ and $S_2 = \sum_i (\bar{X}_i - \bar{X})^2 = 37.91$. A plot of the contour π_x, as a function of the variance components (θ_1, θ_2), is shown in Figure 3; this is based on the new Monte Carlo solution presented in [45]. The dot in the center marks the mode of π_x, the maximum likelihood estimator, $\hat{\theta} = (4.29, 6.15)$.

For estimation of the treatment variance component, Portnoy [51] suggests the following variation on the usual squared-error loss function:

$$\ell_a(\theta) = \frac{(\theta_2 - a)^2}{(\theta_1 + n\theta_2)^2}, \quad a > 0. \tag{14}$$

There are different ways to assess the risk associated with actions relative to this loss function. One is the oracle assessment, another is the fully likelihood-based assessment proposed in [11], the Jeffreys-prior Bayes assessment as in [61], and the IM assessment proposed here. A plot of those assessments is shown in Figure 4. The key observation is that these all have similar shapes and magnitudes, as expected; the IM assessment favors an action smaller than the actions favored by the other assessments, which is consistent with both the loss function and IM solution's inherent reliability/safety properties. How these procedures stack up in terms of their frequentist risk properties will be investigated elsewhere.

4. Invariant decision problems

4.1. Setup. Let $\mathscr{G}$ denote a group of bijections $g : \mathbb{X} \to \mathbb{X}$ acting on $\mathbb{X}$, with function composition $\circ$ as the binary operation. As is customary in the literature, I'll write gx for the image of $x \in \mathbb{X}$ under transformation $g \in \mathscr{G}$; and if g_1 and g_2 are two group elements, then $g_1 \circ g_2$ denotes their composition. Since $\mathscr{G}$ is a group, it's associative, i.e., $g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$ for all $g_1, g_2, g_3 \in \mathscr{G}$, it contains the identity transformation, and for every $g \in \mathscr{G}$, there exists an inverse $g^{-1} \in \mathscr{G}$ such that $g \circ g^{-1} = g^{-1} \circ g = \text{identity}$. Examples include location shifts, rescaling, rotations, and permutations.

The group $\mathscr{G}$ connects to the statistical model as follows. Suppose that, for each $g \in \mathscr{G}$ and each $\theta \in \mathbb{T}$, there exists a corresponding $\bar{g}\theta \in \mathbb{T}$ such that

$$\mathsf{P}_\theta(gX \in \cdot) = \mathsf{P}_{\bar{g}\theta}(X \in \cdot), \quad (\theta, g) \in \mathbb{T} \times \mathscr{G}. \tag{15}$$

For example, if the distribution of X depends on a location parameter θ, then the distribution of $X + \tau$ depends on parameter $\theta + \tau$. When the statistical model $\{\mathsf{P}_\theta : \theta \in \mathbb{T}\}$ satisfies (15), it's called an *invariant statistical model*. It's often the case in applications that the distributions P_θ have densities respect to some underlying dominating measure, and then invariance means

$$p_\theta(x) = p_{\bar{g}\theta}(gx)\,\chi(g), \quad x \in \mathbb{X},\ \theta \in \mathbb{T},\ g \in \mathscr{G}, \tag{16}$$

where $\chi(g)$ is the "multiplier," a change-of-variables Jacobian term. For details, see [53, Ch. 6] and [22].

Define $\overline{\mathscr{G}}$ as the collection of all those $\bar{g} : \mathbb{T} \to \mathbb{T}$, corresponding to the mappings $g \in \mathscr{G}$; this too is a group. A further simplification, commonly used in the literature [e.g., 53, p. 371], is to assume $\mathbb{T} = \mathscr{G} = \overline{\mathscr{G}}$. This assumptions means I don't have to distinguish g, $\bar{g}$, and θ or define functions that connect them.

The set $\mathscr{G}x = \{gx : g \in \mathscr{G}\} \subseteq \mathbb{X}$ is called the *orbit* of $\mathscr{G}$ corresponding to x. The orbits partition $\mathbb{X}$ into equivalence classes, so every point $x \in \mathbb{X}$ falls on exactly one orbit. This partition can be used to construct a new coordinate system on $\mathbb{X}$ which will be useful for us in what follows. Identify $x \in \mathbb{X}$ with (g_x, u_x), where $u_x \in \mathbb{U}$ denotes the label of orbit $\mathscr{G}x$ and $g_x \in \mathscr{G}$ denotes the position of x on the orbit $\mathscr{G}x$.

Model Assumptions. Let $\{p_\theta : \theta \in \mathbb{T}\}$ be a family of densities invariant with respect to a locally compact topological group $\mathscr{G}$ in the sense of (16) and, as explained above, take $\mathbb{T} = \mathscr{G} = \overline{\mathscr{G}}$. In addition, the following hold:

A1. The left Haar measure λ and the corresponding right Haar measure ρ on (the Borel σ-algebra of) $\mathscr{G}$ exist and are unique up to scalar multiples.

A2. There exists a bijection $t : \mathbb{X} \to \mathscr{G} \times \mathbb{U}$, with both t and t^{-1} measurable, that maps $x \in \mathbb{X}$ to its position–orbit coordinates $(g_x, u_x) \in \mathscr{G} \times \mathbb{U}$.

A3. The distribution of $t(X) = (G, U) \in \mathscr{G} \times \mathbb{U}$ induced by the distribution of $X \sim \mathsf{P}_\theta$ has a density with respect to $\lambda \times \mu$ for some measure μ on $\mathbb{U}$.

A few quick remarks are in order. First, for A1, existence and uniqueness of the left and associated right Haar measures on locally compact topological groups is a classical result [e.g., 28, 49]. For A2, note that $t(gx) = (g \circ g_x, u_x)$ for all $g \in \mathscr{G}$. That is, g only acts on the first coordinate in t, so it's invariant with respect to $\mathscr{G}$ in the second coordinate. For A3, existence of a joint density with respect to a product measure simply ensures that there will be no difficulty in defining a conditional distribution for G, given $U = u$. Finally, $U = U_X$, as a function of X, is an ancillary statistic.

The simplest example is a location model where $\mathscr{G} = \overline{\mathscr{G}} = (\mathbb{R}, +)$. This is an abelian group, so the left and right Haar measures are the same and both equal to Lebesgue measure. The function $x \mapsto t(x)$ in A2 consists of two components: in its "g_x" coordinate an equivariant function of x that estimates the location and, in its "u_x" component, an invariant function of x, such as residuals. For example, $g_x = \bar{x}$ the arithmetic mean of $x = (x_1, \ldots, x_n)$ and $u_x = \{x_i - \bar{x} : i = 1, \ldots, n\}$. Note that the u_x coordinate satisfies a constraint, so, after it's represented in a suitable lower-dimensional space $\mathbb{U}$, μ can be taken as Lebesgue measure there. Many other problems fit this general form [e.g., 24, Ch. 1–2, including exercises].

4.2. Bayes–fiducial–IM connection. When the statistical model is invariant with respect to a group of transformations, an interesting connection between the familiar Bayesian/fiducial solution and the IM solution emerges. This connection is through the IM's credal set.

Recall that I'm assuming $\mathbb{T} = \mathscr{G} = \overline{\mathscr{G}}$. In this case, generic values of $\theta \in \mathbb{T}$ can be identified with transformations in $\mathscr{G}$; the same goes for the uncertain variable Θ. So, in what follows, I'll treat θ as a transformation on $\mathbb{X}$ that can be inverted to θ^{-1} and can be composed via $\circ$ with other transformations in $\mathscr{G}$. A key result [53, Cor. 6.64] is that the density of X or, equivalently, of $t(X) = (G, U)$, under P_θ, is given by

$$p_\theta(g, u) = f(\theta^{-1} \circ g, u), \tag{17}$$

where the function $f : \mathscr{G} \times \mathbb{U} \to \mathbb{R}$ doesn't directly depend on θ. The particular form of f isn't important—all that matters is how the right-hand side above depends on θ. Next are two important consequences of (17).

- Lemma 6.65 in [53] says that the Bayesian posterior distribution for Θ, given $x \equiv (g, u)$, has a density with respect to the right Haar prior measure

$$q_x^\star(\theta) = c_g\, c_u'\, f(\theta^{-1} \circ g, u), \tag{18}$$

where c_g and c_u' only depend on the g and u-components of x, respectively. This agrees with the accepted fiducial distribution for Θ, given $x \equiv (g, u)$, in group invariant models.

- Proposition 2 in [42] shows that the relative likelihood is given by $R(x, \theta) = d_u\, f(\theta^{-1} \circ g, u)$, where d_u depends only on the u-component of x.

The importance of these points from the following calculation. Write $\pi_{g|u}(\theta)$ for the possibilistic IM's contour, where the subscript is meant to emphasize that the conditional distribution of $X \equiv (G, U)$, given the observed value of u, is used in the validification step (1). Then

$$\pi_{g|u}(\theta) = \mathsf{P}_\theta\{R(X, \theta) \le R(x, \theta) \mid U = u\}$$

$$
\begin{aligned}
&= \mathsf{P}_\theta\{f(\theta^{-1}\circ G, u) \le f(\theta^{-1}\circ g, u) \mid U = u\} \\
&= \mathsf{P}\{f(H, u) \le f(\theta^{-1}\circ g, u) \mid U = u\} \\
&= \mathsf{Q}_x^\star\{f(H, u) \le f(\theta^{-1}\circ g, u)\} \\
&= \mathsf{Q}_x^\star\{q_x^\star(\Theta) \le q_x^\star(\theta)\},
\end{aligned}
$$

where the first line is by definition; the second line is by simplification; the third line is by the fact that $H := \theta^{-1}\circ G$ is a pivot with respect to the conditional distribution of G, given $U = u$, under P_θ; the fourth line is by the fact that $H := \Theta^{-1}\circ g$ is also a pivot with respect to the Bayes posterior $\Theta \sim \mathsf{Q}_x^\star$ and has the same distribution as H in previous line; and the last line is by simplification via (18).

The last line in the above display can be recognized as the probability-to-possibility transform applied to the Bayesian/fiducial posterior distribution. This observation leads to a new proof of the main result in [42].

Theorem 4.1. *Under the invariant model setup described above, the Bayesian/fiducial posterior distribution $\mathsf{Q}_x^\star$ is the maximal element in the credal set $\mathscr{C}(\overline{\Pi}_{g|u})$ associated with the possibilistic IM. That is, $\mathsf{Q}_x^\star$ satisfies*

$$
\mathsf{Q}_x^\star\big(\{\theta : \pi_{g|u}(\theta) > \alpha\}\big) = 1 - \alpha, \quad \alpha \in [0,1]. \tag{19}
$$

The take-away message is that, under an invariant statistical model, the familiar Bayes/fiducial distribution has a new interpretation as the maximal element in the possibilistic IM's credal set. For example, this explains why these probabilistic solutions offer exact confidence intervals when applied to these specific models.

4.3. Risk assessments. The loss function is invariant if for every $\bar g \in \overline{\mathscr{G}}$ there exists $\tilde g : \mathbb{A} \to \mathbb{A}$ with

$$
\ell_{\tilde g a}(\bar g \theta) = \ell_a(\theta) \quad \text{all } (a, \theta).
$$

The collection $\widetilde{\mathscr{G}}$ of all such maps $\tilde g$ also forms a group under function composition; just as above, for simplicity, I'll assume below that $\mathscr{G} = \overline{\mathscr{G}} = \widetilde{\mathscr{G}} = \mathbb{T}$.

A well-known fact [e.g., 53, Theorem 6.59] is that, when the loss function is invariant, the minimum risk equivariant decision rule is the formal Bayes rule, the minimizer of the expected loss with respect to the posterior distribution based on the right Haar prior ρ. In the present notation, this optimal decision rule is

$$
\hat a(x) = \inf_{a \in \mathbb{A}} \mathsf{Q}_x^\star \ell_a. \tag{20}
$$

Below I show that, under an additional condition on the loss function, the same $\hat a(x)$ minimizes the IM's upper expected loss and, moreover, the IM and Bayes risk assessments of $\hat a$ are identical, i.e., $\overline{\Pi}_x \ell_{\hat a(x)} = \mathsf{Q}_x^\star \ell_{\hat a(x)}$.

The specific condition I assume about the loss function is unfamiliar, but not remarkably uncommon. Define a data-dependent partial order, $\preceq_x$, on $\mathbb{T}$ as follows:

$$
\theta \preceq_x \theta' \iff \pi_x(\theta) \le \pi_x(\theta'). \tag{21}
$$

Then the maximum likelihood estimator is the "$\preceq_x$-largest" element of $\mathbb{T}$ and the more confidence sets $\{C_\alpha(x) : \alpha \in [0,1]\}$ in (4) that an element θ belongs to, the "$\preceq_x$-larger" it is. Then the specific condition I impose is that, for $\hat a(x)$ as defined in (20), the loss function $\theta \mapsto \ell_{\hat a(x)}(\theta)$ is $\preceq_x$-increasing, i.e.,

$$
\theta \preceq_x \theta' \implies \ell_{\hat a(x)}(\theta) \le \ell_{\hat a(x)}(\theta'). \tag{22}
$$

The simplest example of such a loss function is squared error loss in a symmetric location parameter model. A general investigation into the properties of probability distribution families under partial orders like the one above can be found in Bergmann [7].

Theorem 4.2. *Consider an invariant statistical model, with $\overline{\Pi}_x$ the IM's upper probability and the Bayes posterior $\mathsf{Q}_x^\star$ based on the right Haar prior the maximal element in the IM's credal set. If the loss function is invariant and, for the Bayes rule $\hat a(x)$ in (20), satisfies (22) relative to the partial ordering (21), then*

$$
\inf_{a \in \mathbb{A}} \overline{\Pi}_x \ell_a = \overline{\Pi}_x \ell_{\hat a(x)} = \mathsf{Q}_x^\star \ell_{\hat a(x)}. \tag{23}
$$

Therefore, the Bayes rule $\hat a(x)$ is also the IM's minimum upper expected loss action, and the corresponding risk assessments at $\hat a(x)$ are the same.

Proof. By $\preceq_x$-monotonicity in (22), there exists a function $\beta_x(\cdot)$, whose form isn't important, such that

$$
\ell_{\hat a(x)}(\theta) > t \iff \pi_x(\theta) < \beta_x(t), \quad t \ge 0. \tag{24}
$$

Using the basic and definition of the Choquet integral,

$$
\begin{aligned}
\overline{\Pi}_x \ell_{\hat a(x)} &= \int_0^\infty \overline{\Pi}_x(\{\theta : \ell_{\hat a(x)}(\theta) > t\})\, dt \\
&= \int_0^\infty \overline{\Pi}_x(\{\theta : \pi_x(\theta) < \beta_x(t)\})\, dt \\
&= \int_0^\infty \beta_x(t)\, dt \\
&= \int_0^\infty \mathsf{Q}_x^\star(\{\theta : \pi_x(\theta) < \beta_x(t)\})\, dt \\
&= \int_0^\infty \mathsf{Q}_x(\{\theta : \ell_{\hat a(x)}(\theta) > t\})\, dt, \\
&= \mathsf{Q}_x^\star \ell_{\hat a(x)},
\end{aligned}
$$

where the second equality is by (24), the third is by definition of $\overline{\Pi}_x$ as a supremum of π_x, the fourth is by (19), the fifth is by (24), and the sixth by the familiar formula for expectations of non-negative random variables. This proves the second equality in (23). For the first equality, recall that $\overline{\Pi}_x \ell_a \ge \mathsf{Q}_x^\star \ell_a$ for all a. So, if there was another a such that $\overline{\Pi}_x \ell_a < \overline{\Pi}_x \ell_{\hat a(x)}$, then that contradicts the definition (20) of $\hat a(x)$ as the minimizer. □

This result generalizes the conclusion drawn in the example presented in Section 3.3 above. It can also be compared wich the main result in [60] establishing that fiducial methods, which inherently have certain desirable confidence properties, also tend to produce good decision procedures in invariant problems. Theorem 4.2 goes even further in the sense that IMs offer stronger and more comprehensive exact frequentist validity properties—i.e., they're not afflicted by false confidence the way fiducial solutions are—and they yield optimal decision procedures in some cases. The additional $\preceq_x$-monotonicity condition (22) on the loss in Theorem 4.2 complicates this comparison, but this isn't necessary for finding the IM's suggested action; I also have no reason to doubt the overall quality of that procedure when monotonicity, but this deserves further investigation.

4.4. Example. As a quick illustration, consider a directional statistics application, as discussed in [33], where the data are supported on the unit circle. A common model for such data is the so-called *von Mises distribution* with density function $p_\theta(x) \propto \exp\{\varkappa \cos(x - \theta)\}$, where x and θ take values on the unit circle, and $\varkappa > 0$ is a concentration parameter assumed here to be known. This model is invariant with respect to the group $\mathscr{G}$ of rotations, and the corresponding right Haar measure is the uniform distribution on the circle. Martin [42] used this as an illustration of Theorem 4.1 above, showing empirically—using the data in [33, Ex. 1.1]—that the maximal element in the IM's credal set is indeed the Bayes posterior distribution based on the right Haar prior. Here I briefly follow up on that illustration to verify the new result in Theorem 4.2. Using that same data, Figure 5 plots the Bayes risk and the IM's upper risk, as a function of the action a, based on the invariant loss function $\ell_a(\theta) = \cos^{-1}(\langle a, \theta \rangle)$, for a in the unit circle; the plot transforms the unit circle to angles in the interval $[0, 2\pi)$. As Theorem 4.2 predicts, the two curves are not the same, but they agree in terms of both the location and value of their common minimum.

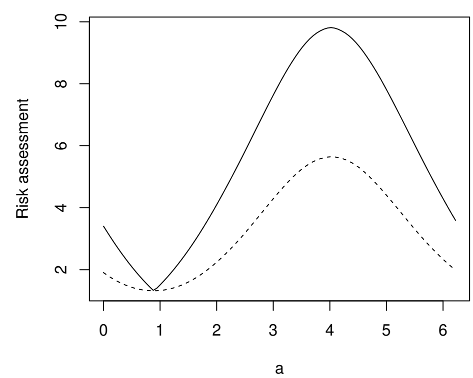

Figure 5. *Plots of the Bayes risk (dashed) and the IM risk (solid) for the directional data illustration in Example 1.1 of [33], where the action is expressed in terms of an angle in the interval* $[0, 2\pi)$*.*

5. Conclusion

To date, IMs have only been used for inference; formal decision-making has not yet been investigated. This paper fills this gap by, first, developing a framework for decision-making under uncertainty based on a possibilistic IM and, second, investigating its decision-theoretic properties. For the first, I follow others and define an upper expected loss via the Choquet integral of the loss with respect to the IM's upper probability, so that different actions can be compared using a natural variation of the minimize-expected-loss principle. For the second, I showed that the IM's validity implies that its assessment, $\overline{\Pi}_X \ell_a$, of an action a would, in a specific sense, not tend to be inconsistent with the quasi-oracle's assessment, thus giving the IM a sense of reliability in the decision-making context which, in turn, provides some protection to the decision-maker. Moreover, the developments in Section 4 reveal that the IM's recommended actions are high-quality, sometimes optimal.

There are a number of directions for future research that can be considered. One is to develop versions of the IM's risk assessment and optimal action for alternatives to preferences based on minimizing upper expected loss, e.g., e-admissibility. Another is to investigate, both theoretically and empirically, the operating characteristics, such as frequentist risk, of the proposed IM-based decision rules. Another is to determine if the monotonicity constraint in Theorem 4.2 can be relaxed/removed.

An important notion in frequentist statistical decision theory is *admissibility*. Unfortunately, the IM's recommended action, defined in (10), is not admissible in general. To see this, consider the d-dimensional normal mean problem. Under loss $\ell_a(\theta) = \|a - \theta\|^2$, the IM's recommended action is $\hat{a}(x) = x$, which corresponds to the least squares, maximum likelihood, fiducial, and flat-prior Bayes estimator, was shown in, e.g., [10, 58] to be inadmissible when $d \geq 3$. This is not especially surprising since validity requires that the IM be "unbiased" in a certain sense which, in the present context, implies that the IM's recommended action be an unbiased estimator in the usual sense. But the familiar bias-variance tradeoff suggests that $r(\cdot; \hat{a})$ can be reduced by choosing a biased $\hat{a}$, e.g., like a proper-prior Bayes rule that's biased and admissible. To introduce the appropriate "bias" into the IM construction, it's natural to consider incorporation of suitable (partial) prior information, e.g., sparsity. A notion of validity under partial prior information is being developed [39, 40], and an extension of the decision-theoretic results presented here to that case, along with admissibility considerations, are part of my ongoing work.

Additional author information

Acknowledgements. Thanks to the anonymous reviewers for their valuable feedback on a previous version of the manuscript. This work is partially supported by the U.S. National Science Foundation, under grants SES–2051225 and DMS–2412628.

References

[1] Thomas Augustin, Gero Walter, and Frank P. A. Coolen. "Statistical inference". In: *Introduction to Imprecise Probabilities*. Wiley Ser. Probab. Stat. Wiley, Chichester, 2014, pp. 135–189.

[2] Michael S. Balch. "Mathematical foundations for a theory of confidence structures". In: *Internat. J. Approx. Reason.* 53.7 (2012), pp. 1003–1019.

[3] Michael S. Balch, Ryan Martin, and Scott Ferson. "Satellite conjunction analysis and the false confidence theorem". In: *Proc. Royal Soc. A* 475.2227 (2019), p. 2018.0565.

[4] James Berger. "The robust Bayesian viewpoint". In: *Robustness of Bayesian Analyses*. Vol. 4. Stud. Bayesian Econometrics. With comments and with a reply by the author. Amsterdam: North-Holland, 1984, pp. 63–144.

[5] James Berger. *Statistical Decision Theory and Bayesian Analysis*. Second. New York: Springer-Verlag, 1985, pp. xvi+617.

[6] James Berger. "Conditioning is the issue". In: *Past, Present, and Future of Statistical Science*. Ed. by Xihong Lin et al. Chapman & Hall/CRC Press, 2014. Chap. 23.

[7] Reinhard Bergmann. "Stochastic orders and their application to a unified approach to various concepts of dependence and association". In: *Stochastic Orders and Decision Under Risk (Hamburg, 1989)*. Vol. 19. IMS Lecture Notes Monogr. Ser. Inst. Math. Statist., Hayward, CA, 1991, pp. 48–73.

[8] Jose-M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester, 1994, pp. xiv+586.

[9] Daniel Bernoulli. "Exposition of a new theory on the measurement of risk". In: *Econometrica* 22 (1954), pp. 23–36.

[10] Lawrence D. Brown. "Admissible estimators, recurrent diffusions, and insoluble boundary value problems". In: *Ann. Math. Statist.* 42 (1971), pp. 855–903.

[11] Marco E. G. V. Cattaneo. "Likelihood decision functions". In: *Electron. J. Stat.* 7 (2013), pp. 2924–2946.

[12] Leonardo Cella and Ryan Martin. "Possibility-theoretic statistical inference offers performance and probativeness assurances". In: *Internat. J. Approx. Reason.* 163 (2023), p. 109060.

[13] Céline Cunen, Nils Lid Hjort, and Tore Schweder. "Confidence in confidence distributions!" In: *Proc. Roy. Soc. A* 476 (2020), p. 20190781.

[14] Arthur P. Dempster. "The Dempster–Shafer calculus for statisticians". In: *Internat. J. Approx. Reason.* 48.2 (2008), pp. 365–377.

[15] Thierry Denœux. "Constructing belief functions from sample data using multinomial confidence regions". In: *Internat. J. of Approx. Reason.* 42.3 (2006), pp. 228–252.

[16] Thierry Denœux. "Likelihood-based belief function: justification and some extensions to low-quality data". In: *Internat. J. Approx. Reason.* 55.7 (2014), pp. 1535–1547.

[17] Thierry Denœux. "Decision-making with belief functions: a review". In: *Internat. J. Approx. Reason.* 109 (2019), pp. 87–110.

[18] Thierry Denœux and Shoumei Li. "Frequency-calibrated belief functions: review and new insights". In: *Internat. J. Approx. Reason.* 92 (2018), pp. 232–254.

[19] Sébastien Destercke and Didier Dubois. "Special cases". In: *Introduction to Imprecise Probabilities*. Wiley Ser. Probab. Stat. Wiley, Chichester, 2014, pp. 79–92.

[20] Didier Dubois. "Possibility theory and statistical reasoning". In: *Comput. Statist. Data Anal.* 51.1 (2006), pp. 47–69.

[21] Didier Dubois, Laurent Foulloy, Gilles Mauris, and Henri Prade. "Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities". In: *Reliab. Comput.* 10.4 (2004), pp. 273–297.

[22] Morris L. Eaton. *Group Invariance Applications in Statistics*. Hayward, CA: Institute of Mathematical Statistics, 1989, pp. vi+133.

[23] Ronald A. Fisher. *Statistical Methods and Scientific Inference*. 3rd. New York: Hafner Press, 1973, pp. viii+182.

[24] Donald A. S. Fraser. *The Structure of Inference*. New York: John Wiley & Sons Inc., 1968, pp. x+344.

[25] Itzhak Gilboa. "Expected utility with purely subjective nonadditive probabilities". In: *J. Math. Econom.* 16.1 (1987), pp. 65–88.

[26] Itzhak Gilboa and David Schmeidler. "Maxmin expected utility with nonunique prior". In: *J. Math. Econom.* 18.2 (1989), pp. 141–153.

[27] Itzhak Gilboa and David Schmeidler. "Additive representations of non-additive measures and the Choquet integral". In: *Ann. Oper. Res.* 52 (1994), pp. 43–65.

[28] Paul R. Halmos. *Measure Theory*. D. Van Nostrand Co., Inc., New York, N. Y., 1950, pp. xi+304.

[29] Jan Hannig, Hari Iyer, Randy C. S. Lai, and Thomas C. M. Lee. "Generalized fiducial inference: a review and new results". In: *J. Amer. Statist. Assoc.* 111.515 (2016), pp. 1346–1361.

[30] Dominik Hose. "Possibilistic Reasoning with Imprecise Probabilities: Statistical Inference and Dynamic Filtering". https://dominikhose.github.io/dissertation/diss_dhose.pdf. PhD thesis. University of Stuttgart, 2022.

[31] Dominik Hose and Michael Hanss. "A universal approach to imprecise probabilities in possibility theory". In: *Internat. J. Approx. Reason.* 133 (2021), pp. 133–158.

[32] Nathan Huntley, Robert Hable, and Matthias C. M. Troffaes. "Decision making". In: *Introduction to Imprecise Probabilities*. Wiley Ser. Probab. Stat. Wiley, Chichester, 2014, pp. 190–206.

[33] Kanti V. Mardia and Peter E. Jupp. *Directional Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2000, pp. xxii+429.

[34] Ryan Martin. "Plausibility functions and exact frequentist inference". In: *J. Amer. Statist. Assoc.* 110.512 (2015), pp. 1552–1561.

[35] Ryan Martin. "On an inferential model construction using generalized associations". In: *J. Statist. Plann. Inference* 195 (2018), pp. 105–115.

[36] Ryan Martin. "False confidence, non-additive beliefs, and valid statistical inference". In: *Internat. J. Approx. Reason.* 113 (2019), pp. 39–73.

[37] Ryan Martin. "An imprecise-probabilistic characterization of frequentist statistical inference". `arXiv:2112.10904`. 2021.

[38] Ryan Martin. "Inferential models and the decision-theoretic implications of the validity property". `arXiv:2112.13247`. 2021.

[39] Ryan Martin. "Valid and efficient imprecise-probabilistic inference with partial priors, I. First results". `arXiv:2203.06703`. 2022.

[40] Ryan Martin. "Valid and efficient imprecise-probabilistic inference with partial priors, II. General framework". `arXiv:2211.14567`. 2022.

[41] Ryan Martin. "Valid and efficient imprecise-probabilistic inference with partial priors, III. Marginalization". `arXiv:2309.13454`. 2023.

[42] Ryan Martin. "Fiducial inference viewed through a possibility-theoretic inferential model lens". In: *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*. Ed. by Enrique Miranda, Ignacio Montes, Erik Quaeghebeur, and Barbara Vantaggi. Vol. 215. Proceedings of Machine Learning Research. PMLR, Nov. 2023, pp. 299–310.

[43] Ryan Martin. "A possibility-theoretic solution to Basu's Bayesian–frequentist via media". In: *Sankhya A* 86 (2024), pp. 43–70.

[44] Ryan Martin. "Which statistical hypotheses are afflicted by false confidence?" In: *BELIEF 2024*. Ed. by Y. Bi, A.-L. Jousselme, and T. Denoeux. Vol. 14909. Lecture Notes in Artificial Intelligence. Switzerland: Springer Nature, 2024, pp. 140–149.

[45] Ryan Martin. "An efficient Monte Carlo method for valid prior-free possibilistic statistical inference". `arXiv:2501.10585`. 2025.

[46] Ryan Martin and Chuanhai Liu. "Inferential models: a framework for prior-free posterior probabilistic inference". In: *J. Amer. Statist. Assoc.* 108.501 (2013), pp. 301–313.

[47] Ryan Martin and Chuanhai Liu. *Inferential Models*. Vol. 147. Monographs on Statistics and Applied Probability. CRC Press, Boca Raton, FL, 2015, pp. xix+254.

[48] Ryan Martin and Jonathan P. Williams. "Asymptotic efficiency of inferential models and a possibilistic Bernstein–von Mises theorem". In: *Internat. J. Approx. Reason.* 180 (2025), Paper No. 109389.

[49] Leopoldo Nachbin. *The Haar Integral*. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London, 1965, pp. xii+156.

[50] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. 2d ed. Princeton University Press, Princeton, N. J., 1947, pp. xviii+641.

[51] Stephen Portnoy. "Formal Bayes estimation with application to a random effects model". In: *Ann. Math. Statist.* 42 (1971), pp. 1379–1402.

[52] Leonard J. Savage. *The Foundations of Statistics*. revised. Dover Publications, Inc., New York, 1972, pp. xv+310.

[53] Mark J. Schervish. *Theory of Statistics*. New York: Springer-Verlag, 1995, pp. xvi+702.

[54] Tore Schweder and Nils Lid Hjort. "Discussion: "Confidence distribution, the frequentist distribution estimator of a parameter: a review"". In: *Int. Stat. Rev.* 81.1 (2013), pp. 56–68.

[55] Teddy Seidenfeld. "A contrast between two decision rules for use with (convex) sets of probabilities: Γ-maximin versus E-admissibility". In: *Synthese* 140.1-2 (2004). With a comment by Isaac Levi, pp. 69–96.

[56] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton, N.J.: Princeton University Press, 1976, pp. xiii+297.

[57] Glenn Shafer. "Belief functions and parametric models". In: *J. Roy. Statist. Soc. Ser. B* 44.3 (1982). With discussion, pp. 322–352.

[58] Charles Stein. "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution". In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*. Berkeley and Los Angeles: University of California Press, 1956, pp. 197–206.

[59] Gunnar Taraldsen. "Joint confidence distributions". https://doi.org/10.13140/RG.2.2.33079.85920. 2021.

[60] Gunnar Taraldsen and Bo Henry Lindqvist. "Fiducial theory and optimal inference". In: *Ann. Statist.* 41.1 (2013), pp. 323–341.

[61] George C. Tiao and Wai-Yuan Tan. "Bayesian analysis of random-effect models in the analysis of variance. I. Posterior distribution of variance-components". In: *Biometrika* 52 (1965), pp. 37–53. DOI: 10.1093/biomet/52.1-2.37. URL: https://doi.org/10.1093/biomet/52.1-2.37.

[62] Matthias C. M. Troffaes. "Decision making under uncertainty using imprecise probabilities". In: *Internat. J. Approx. Reason.* 45.1 (2007), pp. 17–29.

[63] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2014, pp. xviii+415.

[64] Brani Vidakovic. "Γ-minimax: a paradigm for conservative robust Bayesians". In: *Robust Bayesian analysis*. Vol. 152. Lect. Notes Stat. Springer, New York, 2000, pp. 241–259.

[65] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Vol. 42. Monographs on Statistics and Applied Probability. London: Chapman & Hall Ltd., 1991, pp. xii+706.

[66] Rui-Sheng Wang. "Some inequalities and convergence theorems for Choquet integrals". In: *J. Appl. Math. Comput.* 35.1-2 (2011), pp. 305–321.

[67] Larry A. Wasserman. "Belief functions and statistical inference". In: *Canad. J. Statist.* 18.3 (1990), pp. 183–196.

[68] Minge Xie and Kesar Singh. "Confidence distribution, the frequentist distribution estimator of a parameter: a review". In: *Int. Stat. Rev.* 81.1 (2013), pp. 3–39.