

# Fiducial Inference Viewed through a Possibility-Theoretic Inferential Model Lens

Ryan Martin

RGMART13@NCSSU.EDU

Department of Statistics, North Carolina State University, USA

## Abstract

Fisher’s fiducial argument is widely viewed as a failed version of Neyman’s theory of confidence limits. But Fisher’s goal—Bayesian-like probabilistic uncertainty quantification without priors—was more ambitious than Neyman’s, and it’s not out of reach. I’ve recently shown that reliable, prior-free probabilistic uncertainty quantification must be grounded in the theory of imprecise probability, and I’ve put forward a possibility-theoretic solution that achieves it. This has been met with resistance, however, in part due to the statistical community’s singular focus on confidence limits. Indeed, if imprecision isn’t needed to answer confidence-limit-related questions, then what’s the point? In this paper, for a class of practically useful models, I explain specifically why the fiducial argument gives valid confidence limits, i.e., it’s the “best probabilistic approximation” of the possibilistic solution I recently advanced. This sheds new light on what the fiducial argument is doing and on what’s lost in terms of reliability when imprecision is ignored and the fiducial argument is pushed for more than just confidence limits.

**Keywords:** Bayesian, confidence distribution, false confidence, generalized fiducial, group invariance

## 1. Introduction

In the early 20th century, statistical inference and Bayesian inverse probability were synonymous. Fisher, dissatisfied with the Bayesians’ insistence on an *a priori* distribution and their default use of flat prior distributions, put forward a novel alternative—the *fiducial argument* (e.g., Fisher, 1930, 1933, 1935). At a high level, the fiducial argument takes the model-based probabilities for the observable data, depending on fixed parameters, and *flips* them into probability statements about the unknown parameters, depending on the observed data; see, e.g., Zabell (1992), Seidenfeld (1992), and Dawid (2020) for details. Fisher’s proposal was ingenious but too good to be true in general, so, naturally, both supporters and skeptics carefully scrutinized Fisher’s claims, some even attempted reformulating Fisher’s proposal. Most notable is Neyman’s theory of *confidence limits*, which Neyman himself considered to be “an extension of the previous results of Fisher” (Neyman, 1941). Besides Neyman, the work of Dempster (e.g., Dempster, 1967, 1968) aimed at fixing/extending the fiducial argument, and was

a key development in imprecise probability theory. More recently, Hannig and others have developed a theory of *generalized fiducial inference* in which the “fiducial flip” can be applied more systematically; see, e.g., Hannig et al. (2016), Murph et al. (2021), and Section 2.2.

One thing that Fisher and Neyman would’ve agreed on is that the fiducial argument and the theory of confidence intervals are distinct. Still, much of the debate surrounding the two theories focused on whether the fiducial argument produced confidence limits. Fisher said yes, he was later proved wrong, hence “Fisher’s biggest blunder” (Efron, 1998). It’s surprising to me that Fisher didn’t shift the conversation away from confidence limits. Even if the fiducial-limits-are-confidence-limits claim was correct, that was a losing battle: a particular confidence limit construction, even an ingenious one, can’t do better than the “best” confidence limits available in a given application.

Lost in this singular confidence-limit focus is that Fisher’s goal was more ambitious: it aimed to provide Bayesian-like probabilistic uncertainty quantification in the absence of prior information. That goal remains the “most important unresolved problem in statistical inference” (Efron, 2013), but, with a few exceptions (e.g., Taraldsen and Lindqvist, 2013), modern efforts in this direction remain focused solely on confidence-limit related questions. This is still a losing battle, I think, so a different perspective is needed.

Simply put, a full-blown probability distribution isn’t needed to get confidence limits. Instead, the construction of data-dependent probability distributions for inference is motivated by a desire to quantify uncertainty more broadly. In that case, the evaluation of proposed solutions ought to be consistent with these broader objectives, so the metrics I’ve been advocating for take the (data-dependent) probability as the primitive, and require that the “probabilities” this distribution assigns to true/false hypotheses tend to not be small/large. The rationale is that judgments will be made based on the magnitudes of these probabilities, so the above condition implies that these inferences would be reliable or *valid*; see Section 2.3. It’s only after realizing that precise probability distributions are incapable of achieving this kind of reliability that the need for imprecise-probabilistic or *possibilistic* considerations. That’s the motivation behind the possibility-theoretic *inferential model* (IM) framework

that I’ve been advocating for recently (e.g., [Liu and Martin, 2021](#); [Martin, 2021a, 2022b](#)).

Despite the clear differences between fiducial and IMs, there are some superficial similarities. In particular, it’s common for the confidence limits derived from an IM to match those derived from a fiducial-like solution. Some authors (e.g., [Cunen et al., 2020](#); [Cui and Hannig, 2022](#)) have recognized this connection and drawn the conclusion that the IM’s imprecision is unnecessary, that it’s somehow enough to work with fiducial/confidence distributions and, more generally, precise probability theory. These arguments are squarely focused on confidence limits, so they overlook the fact that the “confidence” property satisfied by “confidence distributions” isn’t preserved under the probability calculus. For a theory of “confidence distributions” to do anything more than produce confidence limits, it must have its foundations in the theory of imprecise probability.

I’ve already responded to some of these critiques ([Martin, 2021a](#); [Martin et al., 2021](#)), but I have new insights that will help clear up this confusion about confidence limits versus (imprecise) probabilistic inference. Specifically, my goal here is to give a characterization of the relationship between the fiducial solution—which agrees with the default-prior Bayes solution, among others, in the context I’m considering—and the IM solution. Following some background in Section 2 and a description of the models under consideration (Section 3.1), I show in Section 3.3 that the fiducial solution is the maximal, inner probabilistic approximation to the IM’s possibility measure output. This connection, plus the IM’s validity property, explains why the fiducial argument returns genuine confidence limits—agreeing with the IM’s limits—in this class of problems. It also sheds light on why the IM’s imprecision is needed, i.e., the maximal inner probabilistic approximation of the marginal IM need not be the corresponding marginal fiducial distribution. It’s for this and other reasons (Section 4) that no precise-probabilistic fiducial-like argument can resolve Efron’s “most important unresolved problem.” Finally, in Section 5, I offer a partial answer to a key question: which hypotheses are not afflicted by *false confidence* ([Balch et al., 2019](#); [Martin, 2019](#)) relative to the fiducial solution?

## 2. Background

### 2.1. Problem Setup

Let  $X \in \mathbb{X}$  denote observable data—could be a scalar, a vector, a matrix, a collection scalars, vectors, or matrices, or something else entirely. The standard textbook case of a sample from some population is covered by this general setup, as are many others. Let  $P_\theta$  denote a posited statistical model for the observable  $X$ , depending on a parameter  $\theta \in \mathbb{T}$ . Again, the parameter  $\theta$  is very general, but it’s

typically a scalar or a vector. Of course, the parameter is *unknown* and, as is customary, when I’m referring to the uncertain variable I’ll write it as  $\Theta$ .

The goal is to make inference about  $\Theta$  based on *only* the observed value  $x$  of  $X$  and the model. By “make inference” I mean quantify uncertainty about  $\Theta$ , given  $X = x$ , via a (precise or imprecise) probability distribution supported on  $\mathbb{T}$ . Of course, there are many ways that this can be carried out, and I’ll describe those most relevant to this paper below. First I have to stress that there is *no prior knowledge about  $\Theta$  assumed here*, i.e., the prior about  $\Theta$  is vacuous. I don’t believe a vacuous-prior-knowledge assumption is realistic in most applications ([Martin, 2022a](#)), but this is the typical starting point in the literature. This means that neither ordinary Bayesian inference with a single prior distribution ([Berger, 1985](#); [Bernardo and Smith, 1994](#); [Ghosh et al., 2006](#)) nor generalized Bayesian inference with a proper subset of all prior distributions ([Walley, 1991](#); [Augustin et al., 2014](#)) are viable options to achieve the desired goal. Other solutions, like the ones below, are needed.

### 2.2. Fiducial-Like Constructions

Fisher’s original fiducial argument can be directly applied in only a relatively narrow range of problems. Among those, there’s a fairly general—but still narrow—sub-class of problems amenable to a Fisher-style analysis; these are the *invariant statistical models* discussed in Section 3. But despite Fisher’s failure to formulate a fully general “theory of fiducial inference,” his big idea is so appealing that many others have advanced their own versions. I’ll give details here about only two of these alternatives.

**Default-prior Bayes** The basic idea is to take the posterior distribution to have a density (with respect to Lebesgue measure on  $\mathbb{T}$ ) given by

$$q_x(\theta) \propto L_x(\theta) q(\theta), \quad \theta \in \mathbb{T},$$

where  $L_x$  is the likelihood function based on data  $X = x$ ,  $q$  is some non-negative function, not necessarily a probability density, and the proportionality constant is determined by integrating the expression in the above display. Of course, if that integral diverges, which is possible when  $q$  isn’t a density, then the posterior isn’t well defined. This idea goes back at least to Laplace, who suggested that, in the absence of genuine subjective prior input for  $\Theta$ , one can take a flat, uniform prior distribution on  $\mathbb{T}$ . The uniform prior’s lack of reparametrization invariance was later resolved by [Jeffreys \(1946\)](#). While Jeffreys’s class of priors is widely used, it has two shortcomings: first, the Fisher information it depends on doesn’t always exist (e.g., [Shemyakin, 2014](#); [Lin et al., 2019](#)); second, a “good” default prior for  $\Theta$  can’t support “good” posteriors for all features of  $\Theta$  (e.g., [Fraser, 2011](#);

Fraser et al., 2016). Reference priors (e.g., Berger et al., 2009; Bernardo, 1979) aim to overcome these issues.

**Generalized fiducial** Following Hannig et al. (2016), the starting point is a data-generating equation

$$X = a(\theta, Z), \quad Z \sim P, \quad (1)$$

where  $a$  is a known function and  $P$  is a known probability distribution. This is familiar in the context of data simulation. For inference, this expression effectively links the observable data  $X$  to the uncertain  $\theta$  through a random variable with known distribution. This is the basic starting point for many different frameworks, including those in Fraser (1968), Dawid and Stone (1982), and Dempster (2008). Hannig defines the *generalized fiducial distribution* for  $\theta$ , given  $X = x$ , as the weak limit (as  $\varepsilon \rightarrow 0$ ) of the random variable (function of  $Z$ ) that solves the constrained optimization problem

$$\arg \min_{\theta \in \mathbb{T}} d(x, a(\theta, Z)) \quad \text{s.t.} \quad \min_{\theta \in \mathbb{T}} d(x, a(\theta, Z)) \leq \varepsilon,$$

where  $d$  is an appropriate distance measure. I'll focus here on the case where the above solution is unique, but that's not necessary. When other suitable conditions are met, there's a simple formula for the generalized fiducial density, i.e.,

$$q_x(\theta) \propto L_x(\theta) J(x, \theta), \quad (2)$$

where  $J(x, \theta)$  is basically a Jacobian term resulting from the transformation of  $(Z, P)$  to the solution  $\theta$  of the above optimization problem. Despite the obvious similarities, there's a key difference between Bayes's rule and the right-hand side of (2):  $J$  is not a prior density chosen by the user, it's determined by the model (and the distance  $d$ ).

The two very different approaches described above produce similar looking inferential output. For the class of problems considered in Section 3, their respective outputs are the same. More generally, both return probability distribution with density functions determined by the likelihood times a "weight function." Given their similar forms, one might expect the two solutions to have similar properties. As is common, let's consider large-sample case where  $x^n = (x_1, \dots, x_n)$  is an iid sample of size  $n \rightarrow \infty$  from a common distribution depending on a parameter  $\theta \in \mathbb{T} \subseteq \mathbb{R}^D$ . When standard regularity conditions are satisfied, both the above solutions have a corresponding *Bernstein–von Mises theorem*, which goes as follows. Let  $G_{x^n}$  denote the  $D$ -dimensional Gaussian distribution with mean equal to the maximum likelihood estimator and covariance matrix equal to  $n^{-1}$  times the (observed) Fisher information matrix. Then the theorem states that the total variation distance between  $Q_{X^n}$  and  $G_{X^n}$  is vanishing (in  $P_\theta$ -probability) as  $n \rightarrow \infty$ ; for a proof in the Bayesian

case, see van der Vaart (1998, Ch. 10.2), and details for the generalized fiducial distribution can be found in Hannig et al. (2016) and the references therein.

The total-variation distance is strong enough metric to ensure that confidence limits derived from the Bayes or fiducial distributions are, in fact, confidence limits, at least approximately as  $n \rightarrow \infty$ . For the most part, the theory behind these solutions stops here—because the focus is squarely on confidence limits. As I explained above, a focus on confidence limits will fall short of resolving the "most important unresolved problem." To reach the goal, a broader perspective on uncertainty quantification is needed.

### 2.3. Inferential Models

The *inferential model* (IM) formulation, first developed in Martin and Liu (2013, 2015), has output that takes the mathematical form of a possibility distribution or, equivalently, a consonant belief function. Their motivation for imprecision was that, in order to be reliable when quantifying uncertainty more broadly than with confidence limits, one needs to be more conservative. The *possibilistic* brand of imprecision, compared to other imprecise probability models (e.g., lower previsions), is ideally suited to achieve the error rate control properties that statisticians desire.

More recently, Martin (2022a) developed a simpler and more flexible IM construction, one that makes direct use of the model's likelihood function rather than a data-generating equation like in (1). Let  $L_x(\theta)$  denote the likelihood function, and define the *relative likelihood*

$$R(x, \theta) = \frac{L_x(\theta)}{\sup_{\vartheta \in \mathbb{T}} L_x(\vartheta)}, \quad \theta \in \mathbb{T}. \quad (3)$$

Next, define the *possibility contour*

$$\pi_x(\theta) = P_\theta\{R(X, \theta) \leq R(x, \theta)\}, \quad \theta \in \mathbb{T}. \quad (4)$$

Note that  $\pi_x$  has maximum value 1, which is attained at a maximum likelihood estimator  $\hat{\theta}_x \in \arg \max_{\theta} L_x(\theta)$ . The contour function determines the IM's imprecise-probabilistic output via the formulae

$$\begin{aligned} \overline{\Pi}_x(A) &= \sup_{\theta \in A} \pi_x(\theta) \\ \underline{\Pi}_x(A) &= 1 - \overline{\Pi}_x(A^c), \quad A \subseteq \mathbb{T}. \end{aligned} \quad (5)$$

Here  $\underline{\Pi}_x$  and  $\overline{\Pi}_x$  are conjugate lower and upper probabilities or, more specifically,  $\underline{\Pi}_x$  and  $\overline{\Pi}_x$  are necessity and possibility measures, respectively. I'll call the IM's output *possibilistic* to emphasize that  $\overline{\Pi}_x$  is a possibility measure. Details concerning interpretation of the IM's output can be found in Martin (2022a,b); details on computation can be found in Syring and Martin (2021) and Hose et al. (2022).

The possibility contour expression in (4) is familiar—it's the p-value for a likelihood ratio test of the null hypothesis

$H_0 : \Theta = \theta$ —but that’s not the reason why the contour is defined in this way. The principles behind this construction are presented in [Martin \(2022b, Sec. 4\)](#).

The IM framework originated from the desire to achieve both Bayesian-like “probabilistic” uncertainty quantification and frequentist-style calibration and error rate control guarantees. Indeed, of primary importance is that the IM’s output be *valid* in the sense that  $x \mapsto \overline{\Pi}_x$  satisfy

$$\sup_{\theta \in A} P_\theta \{ \overline{\Pi}_X(A) \leq \alpha \} \leq \alpha, \quad \alpha \in [0, 1], \quad A \subseteq \mathbb{T}. \quad (6)$$

There’s an equivalent condition in terms of  $x \mapsto \overline{\Pi}_x$ , but it’s not needed here. The intuition behind (6) is as follows: assigning a small  $\overline{\Pi}_x$ -value to a true assertion might lead to an erroneous inference, so (6) helps protect the decision-maker by ensuring that such events are sufficiently improbable. This is the IM’s reliability guarantee. It’s easy to see that the IM with upper probability  $\overline{\Pi}_x$  in (5) satisfies (6): it follows immediately from the inequality  $\overline{\Pi}_X(A) \geq \pi_X(\Theta)$ , for all  $A$  that contain  $\Theta$ , and the fact that  $\pi_X(\Theta)$  is stochastically no smaller than  $\text{Unif}(0, 1)$ . Note that (6) is different from the usual frequentist Type I error control in testing. Validity covers all  $A$ ’s at once, hence ensuring reliability even when interest is in a (potentially non-linear) feature  $\psi(\Theta)$  of  $\Theta$ ; just take, e.g.,  $A = \{ \theta : \psi(\theta) \leq 7 \}$ .

Finally, adjustments can be made to the IM construction above that do not affect the validity and, in fact, can often help to improve its efficiency. One is that it suffices to work with a minimal sufficient statistic. Also, for cases where the minimal sufficient statistic has an ancillary component, like in the class of problems considered below, I showed ([Martin, 2022b, Sec. 6](#)) that the IM construction should be carried out conditional on the ancillary statistic. Further details on this will be given in Section 3.2.

### 3. A Fiducial–IM Connection

#### 3.1. Invariant Statistical Models

Let  $\mathcal{G}$  denote a group of bijections  $g : \mathbb{X} \rightarrow \mathbb{X}$  acting on  $\mathbb{X}$ , with function composition  $\circ$  as the binary operation.<sup>1</sup> As is customary in the literature, I’ll write  $gx$  for the image of  $x \in \mathbb{X}$  under transformation  $g \in \mathcal{G}$ ; and if  $g_1$  and  $g_2$  are two group elements, then  $g_1 \circ g_2$  denotes their composition. Since  $\mathcal{G}$  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  $g_1, g_2, g_3 \in \mathcal{G}$ , it contains the identity transformation  $e$ , and for every  $g \in \mathcal{G}$ , there exists an inverse  $g^{-1} \in \mathcal{G}$  such that  $g \circ g^{-1} = g^{-1} \circ g = e$ . Some examples include location shifts, rescaling, rotations, affine transformations, and permutations. Note that the group  $\mathcal{G}$  is typically finite-dimensional, e.g., rotations in Euclidean space correspond to unit-determinant matrices.

<sup>1</sup>If  $\mathbb{X}$  is a product space, then  $\mathcal{G}$  can be extended, if necessary, by applying it coordinate-wise.

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

$$P_\theta(gX \in \cdot) = P_{\bar{g}\theta}(X \in \cdot), \quad (\theta, g) \in \mathbb{T} \times \mathcal{G}. \quad (7)$$

The most common example of this is where the distribution of  $X$  depends on a location parameter  $\theta$  and, consequently, the distribution of  $X + a$  depends on the location parameter  $\theta + a$ . When the statistical model  $\{P_\theta : \theta \in \mathbb{T}\}$  satisfies (7), it’s called an *invariant statistical model*. A good textbook presentation on this is [Schervish \(1995, Ch. 6\)](#); a more comprehensive account is given in [Eaton \(1989\)](#). Here I’ll present only the necessary details.

Define  $\overline{\mathcal{G}}$  as the collection of those bijections  $\bar{g} : \mathbb{T} \rightarrow \mathbb{T}$ , corresponding to the mappings  $g \in \mathcal{G}$ . It is easy to check that  $\overline{\mathcal{G}}$  is a group. As is often the case in applications, I’ll assume that the distributions  $P_\theta$  all have a density respect to some underlying  $\sigma$ -finite measure. So I’ll follow the suggestion<sup>2</sup> from Theorem 3.1 in [Eaton \(1989\)](#) and say that the family  $\{p_\theta : \theta \in \mathbb{T}\}$  is invariant with respect to  $\mathcal{G}$  if

$$p_\theta(x) = p_{\bar{g}\theta}(gx) \chi(g), \quad x \in \mathbb{X}, \theta \in \mathbb{T}, g \in \mathcal{G}, \quad (8)$$

where  $\chi(g)$  is the “multiplier” from a change-of-variables. These model assumptions aren’t necessary to define the fiducial, Bayes, and IM solutions; this just sets a context in which a connection between them can be made.

The set  $\mathcal{G}x = \{gx : g \in \mathcal{G}\} \subseteq \mathbb{X}$  is called the *orbit* of  $\mathcal{G}$  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 and useful coordinate system on  $\mathbb{X}$ . Identify  $x \in \mathbb{X}$  with  $(g_x, u_x)$ , where  $u_x \in \mathbb{U}$  denotes the label of orbit  $\mathcal{G}x$  and  $g_x \in \mathcal{G}$  denotes the position<sup>3</sup> of  $x$  on the orbit  $\mathcal{G}x$ .

Henceforth, I’m going to sacrifice a bit of generality for the sake of readability. As is common in the literature, instead of assuming that  $\mathcal{G}$  and  $\overline{\mathcal{G}}$  are isomorphic, I’m going to follow [Schervish \(1995, p. 371\)](#) and assume that  $\mathbb{T} = \mathcal{G} = \overline{\mathcal{G}}$ ; this means we don’t have to distinguish  $g, \bar{g}$ , and  $\theta$  and we don’t have to track functions that connect the three. With this in mind, let’s agree on the following

**Model Assumptions** Let  $\{p_\theta : \theta \in \mathbb{T}\}$  be a family of densities invariant with respect to a locally compact topological group  $\mathcal{G}$  in the sense of (8) and, as explained above, take  $\mathbb{T} = \mathcal{G} = \overline{\mathcal{G}}$ . In addition, the following hold:

- A1. The left Haar measure  $\lambda$  and the corresponding right Haar measure  $\rho$  on (the Borel  $\sigma$ -algebra of)  $\mathcal{G}$  exist and are unique up to scalar multiples.

<sup>2</sup>Eaton’s Theorem 3.1 says that (8) implies (7) but, the converse is only “almost” true. That is, (7) and existence of densities implies (8) on a null set that can depend on  $(\theta, g)$ . However, this null set dependence on  $(\theta, g)$  can be avoided in all the applications that he’s familiar with.

<sup>3</sup>Position is relative to a predetermined reference point on the orbit: if  $r_x$  is that reference point on  $\mathcal{G}x$ , then  $g_x$  is defined such that  $x = g_x r_x$ .

- A2. There exists a bijection  $t : \mathbb{X} \rightarrow \mathcal{G} \times \mathbb{U}$ , with both  $t$  and  $t^{-1}$  measurable, that maps  $x \in \mathbb{X}$  to its position–orbit coordinates  $(g_x, u_x) \in \mathcal{G} \times \mathbb{U}$ .
- A3. The distribution of  $t(X) = (G, U) \in \mathcal{G} \times \mathbb{U}$  induced by the distribution of  $X \sim P_{X|\theta}$  has a density with respect to  $\lambda \times \mu$  for some measure  $\mu$  on  $\mathbb{U}$ .

A few quick remarks about these assumptions are in order. First, it’s not necessary that  $\mathbb{T} = \mathcal{G} = \overline{\mathcal{G}}$ , only that  $\mathcal{G}$  and  $\overline{\mathcal{G}}$  are isomorphic and that there’s a bijection that relates  $\mathbb{T}$  and  $\overline{\mathcal{G}}$ . Next, for A1, existence and uniqueness of the left and associated right Haar measures on locally compact topological groups is a classical result and I’ll refer the reader to the corresponding classical texts: [Halmos \(1950\)](#) and [Nachbin \(1965\)](#). For A2, note that  $t(gx) = (g \circ g_x, u_x)$  for all  $g \in \mathcal{G}$ . That is,  $g$  only acts on the first coordinate in  $t$ , so it’s invariant with respect to  $\mathcal{G}$  in the second coordinate—the orbit label is unaffected by actions in  $\mathcal{G}$ . Finally, for A3, existence of a joint density with respect to a product measure simply ensures no difficulty in defining a conditional distribution for  $G$ , given  $U = u$ ; moreover,  $U = U_X$ , as a function of  $X \sim P_\theta$  is an ancillary statistic.

The simplest example is that of a location parameter where  $\mathcal{G} = \overline{\mathcal{G}} = (\mathbb{R}, +)$ . Since this group is abelian, 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  $x = (x_1, \dots, x_n)$  and  $u_x = \{x_i - \bar{x} : i = 1, \dots, n\}$ . Note that the  $u_x$  coordinate satisfies a constraint, so, after it’s represented in a suitable lower-dimensional space  $\mathbb{U}$ ,  $\mu$  can be taken as Lebesgue measure there. There are many other problems that fit this general form; see, e.g., Section 3.4 below and Chapters 1–2 of [Fraser \(1968\)](#), including the exercises.

### 3.2. Fiducial and IM Solutions

For the above-described invariant statistical models, there is a standard/accepted fiducial distribution construction, which I’ll describe below. The IM solution as presented in Section 2.3 is relatively straightforward in this case too.

Just a quick remark on notation before getting started. Recall that, for simplicity, I’m assuming  $\mathbb{T} = \mathcal{G} = \overline{\mathcal{G}}$ . In this case, generic values of  $\theta \in \mathbb{T}$  can be identified with transformations in  $\mathcal{G}$ ; the same goes for the uncertain variable  $\Theta$ . So, in what follows, I’ll treat  $\theta$  as a transformation that maps  $x$  to  $\theta x$ , can be inverted to  $\theta^{-1}$ , and can be composed via  $\circ$  with other transformations in  $\mathcal{G}$ .

A key result (e.g., [Schervish, 1995](#), Corollary 6.64) says that the density of  $X$  or, equivalently, the density of  $t(X) = (G, U) \in \mathcal{G} \times \mathbb{U}$ , under  $P_\theta$ , is given by

$$p_\theta(g, u) = f(\theta^{-1} \circ g, u), \quad (\lambda \times \mu)\text{-almost all } (g, u), \quad (9)$$

where  $f : \mathcal{G} \times \mathbb{U} \rightarrow \mathbb{R}$  is a fixed function that doesn’t directly depend on  $\theta$ . The particular form of  $f$  isn’t important for us here—all that matters is how it depends on  $\theta$ .

There are two important consequences of (9). The first is presented as Lemma 6.65 in [Schervish \(1995\)](#).

**Proposition 1** *The fiducial distribution  $Q_x$  of  $\Theta$ , given  $X = x$  or, equivalently, given  $t(X) = (g, u)$ , has a density  $q_x$  with respect to right Haar measure on  $\mathcal{G}$  given by*

$$q_x(\theta) = c_g p_\theta(g | u), \quad \theta \in \mathbb{T}, \quad t(x) = (g, u), \quad (10)$$

where  $c_g$  is a constant that depends only on  $g$ , and  $p_\theta(g | u)$  is the conditional density of  $G$ , given  $U = u$ , derived from the joint density in (9). Also,  $q_x$  in (10) is the Bayesian posterior density under the right Haar prior.

Here’s what’s going on. If  $G$  is as in  $t(X) = (G, U)$ , where  $X \sim P_\theta$ , then the conditional distribution of  $H := \theta^{-1} \circ G$  in  $\mathcal{G}$ , given  $U = u$ , doesn’t depend on  $\theta$ . Therefore,  $H$  is a *pivot*, a function of  $(X, \theta)$  having distribution free of  $\theta$ . Then the fiducial distribution for  $\Theta$ , given  $X = x$  or, equivalently, given  $(G, U) = (g, u)$ , is

$$Q_x(A) = P(g \circ H^{-1} \in A | U = u), \quad A \subseteq \mathbb{T},$$

where the probability on the right-hand side is with respect to the conditional distribution of  $H$ , given  $U = u$ , derived from  $P_\theta$ , which doesn’t depend on  $\theta$ . If that conditional distribution is known, then the fiducial probabilities can be directly evaluated or approximated via Monte Carlo. The Bayes connection can be of some practical benefit, instead of working with the  $(H | U = u)$  conditional distribution, one can apply any of the Monte Carlo methods commonly used to approximate Bayesian posterior distributions.

The next result, which is related to the ideas in the previous paragraph, is the second important consequence of (9), relevant to the IM solution.

**Proposition 2** *The relative likelihood function  $R(x, \theta)$  in (3) depends on  $(\theta^{-1} \circ g, u)$  only, where  $t(x) = (g, u)$ . In particular, if  $t(x) = (g, u)$ , then*

$$R(x, \theta) = d_u f(\theta^{-1} \circ g, u),$$

where  $d_u$  is a constant that depends only on  $u$ .

**Proof** Since  $x \mapsto t(x)$  is a bijection, by (9) we have

$$R(x, \theta) = \frac{p_\theta(g, u)}{\sup_{\vartheta \in \mathbb{T}} p_\vartheta(g, u)} = \frac{f(\theta^{-1} \circ g, u)}{\sup_{\vartheta \in \mathbb{T}} f(\vartheta^{-1} \circ g, u)}.$$

Since  $\vartheta$  is free to vary across all of  $\mathbb{T}$  or, equivalently, all of  $\mathcal{G}$ , so too is  $\vartheta^{-1} \circ g$ . This implies that the supremum on the right-hand side doesn’t depend on  $g$ . The claim follows by taking  $1/d_u$  equal to that supremum. ■

The key observation is that, since  $R(X, \theta)$  depends on  $X$  only through  $(\theta^{-1} \circ G, U)$ , and  $H := \theta^{-1} \circ G$  is a pivot, it follows that  $R(X, \theta)$  is a pivot too. The relative likelihood is a pivot (or at least approximately so, via Wilks’s theorem) in many of the familiar textbook problems, so this is not a surprising result. In any case, it’s now straightforward to write down the IM solution’s contour function:

$$\pi_x(\theta) = \mathbb{P}\{f(H, u) \leq f(\theta^{-1} \circ g, u) \mid U = u\}, \quad \theta \in \mathbb{T},$$

where the probability on the right is the conditional distribution of  $H$ , given  $U = u$ , derived from  $\mathbb{P}_\theta$ , which doesn’t depend on  $\theta$ . Analogous to the formulation in Section 2.3, I can let  $\overline{\Pi}_x(A) = \sup_{\theta \in A} \pi_x(\theta)$  define the IM’s possibility measure output. Where it’s helpful, I’ll write “ $\pi_{g|u}$ ” and “ $\overline{\Pi}_{g|u}$ ” to emphasize that the IM output depends on the observed  $x$  only through  $t(x) = (g, u)$ , and that  $u$  is fixed at its observed value. With this notation, the contour satisfies

$$\mathbb{P}_\theta\{\pi_{G|u}(\theta) \leq \alpha \mid U = u\} \leq \alpha, \quad \alpha \in [0, 1],$$

where  $(G, U)$  on the right-hand side depend implicitly on  $X \sim \mathbb{P}_\theta$ . From here we get the following conditional version of the validity property (6): for each  $A \subseteq \mathbb{T}$ ,

$$\sup_{\theta \in A} \mathbb{P}_\theta\{\overline{\Pi}_{G|u}(A) \leq \alpha \mid U = u\} \leq \alpha, \quad \alpha \in [0, 1].$$

### 3.3. Connection

Roughly speaking, the main result here says that, in the invariant statistical model setting, the fiducial distribution  $\mathbb{Q}_x$  described above is not only a member of the IM’s credal set  $\mathcal{C}(\overline{\Pi}_x)$  but is a “maximal” such member in the sense of being most diffuse, having the heaviest tails, etc. First I need to define this notion of maximality precisely. This will be defined in terms of a generic uncertain variable  $Y$ , taking values in a space  $\mathbb{Y}$ , about which uncertainty is quantified via a possibility measure  $\Pi$ . Recall that the credal set  $\mathcal{C}(\Pi)$  for a possibility measure  $\Pi$  supported on  $\mathbb{Y}$  is the set of all probability measures it dominates, i.e.,  $\mathcal{C}(\Pi) = \{\mathbb{P} \in \text{prob}(\mathbb{Y}) : \mathbb{P}(\cdot) \leq \Pi(\cdot)\}$ .

**Definition 3** *Given a possibility measure  $\Pi$  on a space  $\mathbb{Y}$  with contour  $\pi$ , a probability measure  $\mathbb{P}$  supported on  $\mathbb{Y}$  is a maximal element in the credal set  $\mathcal{C}(\Pi)$  if*

$$\mathbb{P}\{\pi(Y) \leq \alpha\} = \alpha, \quad \text{for all } \alpha \in [0, 1].$$

The well-known characterization (e.g., [Couso et al., 2001](#); [Destercke and Dubois, 2014](#)) says

$$\mathbb{P} \in \mathcal{C}(\Pi) \iff \mathbb{P}\{\pi(Y) \leq \alpha\} \leq \alpha, \quad \alpha \in [0, 1],$$

i.e., if the  $\mathbb{P}$ -probability assigned to the  $\alpha$ -lower level sets of  $\pi$  is no more than  $\alpha$  for every  $\alpha \in [0, 1]$ . So a maximal

element of the credal set assigns exactly the maximal probability to these lower level sets. Alternatively, maximality implies that  $\pi$  characterizes the tails of  $\mathbb{P}$ :

$$\pi(y) = \mathbb{P}\{\pi(Y) \leq \pi(y)\}, \quad y \in \mathbb{Y}.$$

The following result establishes an apparently new connection between the fiducial solution and the possibilistic IM solution in the invariant statistical model setting. Specifically, the fiducial distribution  $\mathbb{Q}_x$  is a maximal member of the IM’s credal set  $\mathcal{C}(\overline{\Pi}_x)$ . That is, for each  $x \in \mathbb{X}$ ,

$$\mathbb{Q}_x\{\pi_x(\theta) \leq \alpha\} = \alpha, \quad \text{for all } \alpha \in [0, 1]. \quad (11)$$

Note that  $\mathbb{Q}_x$  and  $\pi_x$  depend on  $x$  only through  $t(x) = (g, u)$ . Also, for the models under consideration here, the fiducial, generalized fiducial, and default-prior Bayes solutions all agree, so the following result applies to all of these.

**Theorem 4** *Under the setup described in Sections 3.1–3.2, for each fixed  $x$ , the fiducial distribution  $\mathbb{Q}_x$  is a maximal member of the IM’s credal set  $\mathcal{C}(\overline{\Pi}_x)$ .*

**Proof** The fiducial  $\mathbb{Q}_x$ -probability can be rewritten as

$$\mathbb{Q}_x\{\pi_x(\theta) \leq \alpha\} = \mathbb{P}\{\pi_{g|u}(g \circ H^{-1}) \leq \alpha \mid U = u\}.$$

Let  $z_\alpha(u)$  denote the  $\alpha$ -quantile of the conditional distribution of  $f(H, u)$ , given  $U = u$ ; this doesn’t depend on the unknown parameter. Since  $\pi_{g|u}(g \circ h^{-1}) \leq \alpha$  if and only if  $f((g \circ h^{-1})^{-1} \circ g, u) \equiv f(h, u) \leq z_\alpha(u)$ , it follows that

$$\begin{aligned} \mathbb{Q}_x\{\pi_x(\theta) \leq \alpha\} &= \mathbb{P}\{\pi_{g|u}(g \circ H^{-1}) \leq \alpha \mid U = u\} \\ &= \mathbb{P}\{f(H, u) \leq z_\alpha(u) \mid U = u\} \\ &= \alpha, \end{aligned}$$

which proves the claim (11). ■

Theorem 4 shows that, for invariant statistical models, the standard fiducial solution (which agrees with the default-prior Bayesian solution, among others) is the best probabilistic approximation to the IM’s possibilistic output. This eliminates any mystery surrounding the fiducial solution: it’s the maximal inner probabilistic approximation to the valid IM’s possibilistic output. That is, the fiducial argument isn’t magically solving the problem of prior-free probabilistic inference within the theory of precise probability, it’s just the “best one can do” within that theory. Common sense suggests that there’s no free lunch and, indeed, the cost of artificially introducing precision is that one must limit the scope of questions that can be reliably answered; see Section 5 below. It’s true that the IM’s possibilistic output lacks the granularity of probabilistic output, but my claim is that, on its own, without the assistance of prior information, data can only reliably support possibilistic reasoning.

Pushing data any further than this, as fiducial-like solutions do, creates a risk of inferences being unreliable.

There are further conceptual insights coming from Theorem 4. As I mentioned in Section 2.1, the starting point is that there are no non-trivial, *a priori* probability statements that can be made about  $\Theta$ . So, whether it be via the fiducial argument or something else, the  $Q_x$ -probabilities come out of thin air. In the invariant model case, one might be able to justify these based on there being a consensus—most/all of the available approaches identify the same  $Q_x$ . But this doesn't explain what the probabilities *mean*. They're clearly neither subjective nor objective and, to my knowledge, there's no agreed-upon interpretation. This isn't a surprise in light of Theorem 4. Indeed, what's driving the fiducial solution is a weaker, possibilistic uncertainty quantification. The fiducial probability is just one of the many compatible with the IM solution's possibility measure and, therefore, can't have any genuine meaning.

### 3.4. Directional Data Example

Consider direction measurements on the plane, which can be represented by angles relative to a reference point. Real-world examples include wind direction and animal movement studies. More generally, direction measurements can be represented by points on the surface of a hypersphere. This is common in astronomy, where the position of planets, stars, etc. can be described by points on the celestial sphere.

As a very simple illustration, I'll consider the roulette wheel data from Example 1.1 in [Mardia and Jupp \(2000\)](#). The experiment proceeds by spinning a roulette wheel and recording the angle (in radians) of the position at which the wheel stops. Let  $Y^n = (Y_1, \dots, Y_n)$  be the observable angles for  $n$  independent spins. The actual data in this case, which consists of  $n = 9$  angles, is presented graphically in Figure 1. The model under consideration here is the *von Mises distribution* which has a density function

$$p_\theta(y) \propto \exp\{\kappa \cos(y - \theta)\}, \quad y, \theta \in [0, 2\pi).$$

Here  $\kappa > 0$  is a known concentration parameter, taken to be equal to 2 in this illustration. It's no trouble allowing  $\kappa$  to be unknown as well, but that takes us outside of the invariant statistical model context. The parameter  $\theta$  is an unknown mean angle to be inferred from the data.

It's easy to see that the minimal sufficient statistic for this model is the pair

$$X = (\bar{C}, \bar{S}) = (n^{-1} \sum_{i=1}^n \cos Y_i, n^{-1} \sum_{i=1}^n \sin Y_i).$$

Then  $X$  can be treated as “the data.” Since  $(C_i, S_i) = (\cos Y_i, \sin Y_i)$  denotes the Cartesian coordinates of the point on the unit circle corresponding to the angle  $Y_i$ , it's clear that  $X = (\bar{C}, \bar{S})$  is generally inside the unit circle; see Figure 1. The more “concentrated” the  $Y_i$ 's are around a

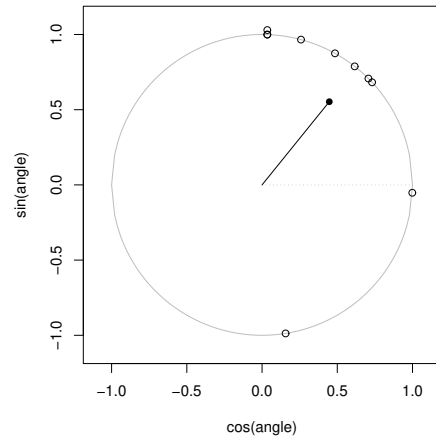


Figure 1: Open circles are the Cartesian coordinates corresponding to the observed angles; two duplicate observations near the north pole are stacked. The solid circle is the arithmetic mean of those Cartesian coordinates, and the angle the black line makes with the dotted reference line is the maximum likelihood estimator  $G$  of  $\theta$ ; the black line's length is the sample concentration  $U$ .

particular angle, the closer  $X$  will be to the circle's boundary. Convert this average position to polar coordinates:

$$G = \arctan(\bar{S}/\bar{C}) \quad \text{and} \quad U = (\bar{C}^2 + \bar{S}^2)^{1/2}.$$

Here  $G$  is the angle the vector  $(\bar{C}, \bar{S})$  makes with the horizontal axis and  $U$  is the Euclidean length of  $(\bar{C}, \bar{S})$ . It's clear that  $(G, U)$  is a bijection of the minimal sufficient statistic  $X$ , so write  $t(X) = (G, U)$ . There is a group  $\mathcal{G}$  consisting of rotations of points in the plane; technically,  $\mathcal{G}$  is the special orthogonal group—orthogonal matrices with unit determinant. Then  $G$  can be interpreted as both an angle or as a rotation in  $\mathcal{G}$  by that angle. Similarly, the interior of the unit circle can be partitioned into distinct concentric circles, which correspond to orbits, and  $U$  determines which of these orbits  $X$  sits on.

The details here align with the general setup above, so I can proceed to carry out the fiducial (or default-prior Bayes) and IM analyses. It's a standard result (e.g., [Mardia and Jupp, 2000](#), Eq. 4.5.5) in the directional statistics literature that the joint density for  $(G, U)$ , factors as

$$p_\theta(g, u) = p_\theta(g | u) p(u) \propto \exp\{\kappa u \cos(g - \theta)\} p(u),$$

where  $p(u)$  denotes the marginal density for  $U$ , which doesn't depend on  $\theta$  and isn't relevant here. The conditional distribution of  $G$ , given  $U = u$ , is easily seen to be von Mises with concentration parameter  $\kappa u$ . The influence

of conditioning on  $U$  makes intuitive sense because, as indicated above, the statistic  $U$  acts like a measure of how “concentrated” the observed angles are around a common angle. It’s the conditional density above that drives both the fiducial and IM analyses. Indeed, the fiducial density is

$$q_x(\theta) \propto \exp\{\kappa u \cos(g - \theta)\}, \quad \theta \in (0, 2\pi],$$

and, after simplification, the IM’s possibility contour is

$$\pi_{g|u}(\theta) = P\{\cos H \leq \cos(g - \theta) \mid U = u\}, \quad \theta \in [0, 2\pi),$$

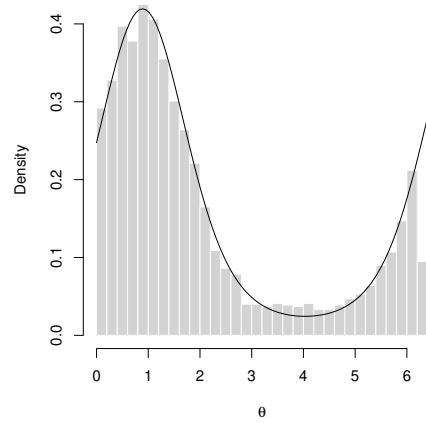
where the probability is with respect to the conditional distribution of  $H := G - \Theta$ , given  $U = u$ , which is simply a von Mises distribution with mean angle 0 and concentration parameter  $\kappa u$ . Plots of  $\theta \mapsto q_x(\theta)$  (overlaid on Monte Carlo samples from  $q_x$ ) and  $\theta \mapsto \pi_{g|u}(\theta)$  are shown in Figure 2. Both point to values of  $\Theta$  near the maximum likelihood estimator  $g = 0.89$  as most plausible, as expected.

#### 4. A New Fiducial Argument?

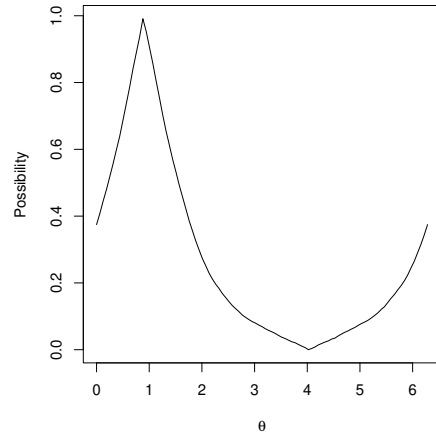
Theorem 4 says that, for a certain class of models, nothing changes if we *define* the fiducial solution as the maximal element in the IM’s credal set. This alternative view the fiducial argument is appealing for at least one reason.

Whether it’s in the invariant statistical model setting or not, we can always *define* the fiducial solution to be the maximal element in the IM’s credal set; see Martin (2021b, Sec. 3) and Taraldsen (2021, Sec. 2). That is, instead of constructing a data-dependent distribution  $Q_x$  and hoping that it satisfies certain properties (e.g., a Bernstein–von Mises theorem), just define the fiducial distribution to be the best probabilistic approximation of the IM’s possibilistic output. This would ensure that the fiducial distribution’s credible sets are genuine confidence sets, which is what “confidence distributions” aim to achieve (e.g., Thornton and Xie, 2020; Nadarajah et al., 2015; Xie and Singh, 2013). The challenge with the suggested strategy is actually finding the best probabilistic approximation. Outside the invariant case, the currently-available fiducial solutions likely don’t correspond to members of the IM’s credal set. So identifying and numerically evaluating the “fiducial distribution” I just defined could be a challenge. One idea is to apply the possibility-to-probability transform in Dubois et al. (2004, Sec. 3.2) to get a probability  $Q_x$  from the IM’s possibility measure  $\overline{\Pi}_x$ , but this is non-trivial even for scalar  $\Theta$ .

To be clear: *I’m not advocating for any fiducial argument.* The validity property described in Section 2.3 is essential to the logic of statistical inference and, as shown in Balch et al. (2019), there are no data-dependent probability distributions that can achieve it. My goal is simply to better understand the fiducial argument. It’s now clear that, at least in the class of problems considered above, the fiducial



(a) Fiducial density & samples



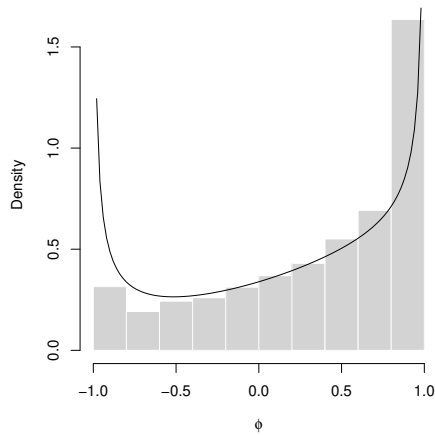
(b) IM possibility contour

Figure 2: Fiducial and IM output for inference on the angle  $\Theta$  based on the roulette wheel data in Figure 1.

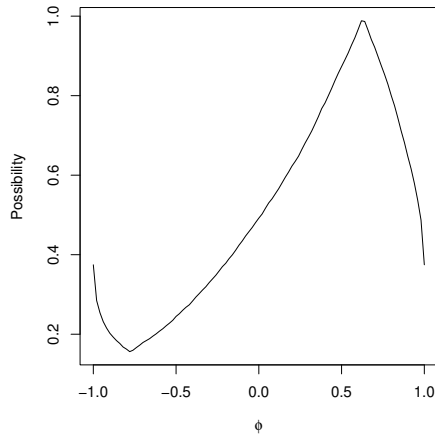
solution corresponds to a maximal probabilistic approximation of the IM’s possibilistic solution. The level sets of the IM’s possibility contour are confidence regions and the connection established in Theorem 4 implies that the fiducial confidence regions agree with these. However, *the connection between fiducial and IMs doesn’t imply that the former is valid like the latter.* Fiducial-like solutions are still at risk for false confidence. This is because the maximal probabilistic approximation of a marginal IM generally isn’t the corresponding marginal fiducial distribution.

For example, Figure 3 shows the marginal fiducial (and default-prior Bayes posterior) distribution and marginal IM possibility contour for  $\Phi = \cos \Theta$  from the example in Section 3.4. Despite the similarities in Figure 2, the (non-linear) marginalization has changed their shapes. Here the fiducial distribution’s mode is pushed close to the boundary,





(a) Marginal fiducial density &amp; samples



(b) Marginal IM possibility contour

 Figure 3: Marginal fiducial distribution and marginal IM possibility contour for  $\Phi = \cos \Theta$  derived from those for  $\Theta$  as displayed in Figure 2.

whereas the IM's mode is at the maximum likelihood estimator  $\cos 0.89 = 0.63$ ; since the modes differ, the fiducial distribution is no longer the best probabilistic approximation of the IM's possibility measure. This post-marginalization discrepancy is a result of differences between the probability and possibility calculi. The possibility calculus is guaranteed to preserve the IM's validity property, whereas the probability calculus offers no such guarantees.

## 5. False Confidence Redux

Balch et al. (2019) show that data-dependent probabilities used for inference suffer from *false confidence*. That is, if  $x \mapsto Q_x$  is such a method, then the false confidence theorem says, for any  $\alpha \in (0, 1)$  and any  $\theta \in \mathbb{T}$ , there exists

hypotheses  $A \subseteq \mathbb{T}$  such that

$$A \ni \theta \quad \text{and} \quad \mathbb{P}_\theta\{Q_X(A) \leq \alpha\} > \alpha. \quad (12)$$

Intuitively, if  $\alpha$  represents the cutoff between “small” and “not small,” then (12) says that the event  $\{Q_X(A) \text{ is small}\}$  has not-small  $\mathbb{P}_\theta$ -probability when  $\theta \in A$ . Since we'd be inclined to doubt the truthfulness of  $A$  when  $Q_X(A)$  is small, the property in (12) creates a risk of, e.g., regularly rejecting true hypotheses. Consequently, inference based on  $Q_X$  is at risk of being systematically misleading.

The false confidence theorem, in its current form, is only an existence result. It could be that the only problematic  $A$ 's are trivial, not of practical interest. Despite the motivating example in Balch et al. (2019) involving non-trivial hypotheses, most statisticians have dismissed this result; apparently they don't believe non-trivial hypotheses can be afflicted. So, the burden is on me/us to push this result to the point that the conclusion can't be denied.

Fortunately, the structure provided by the invariant statistical model allows for a more in-depth investigation into this question. Unfortunately, what I present below still falls short of characterizing those  $A$  that are afflicted with false confidence. I can only give a (fairly general) sufficient condition for a hypothesis  $A$  to be free of false confidence. The necessary conditions that would completely settle the matter are still out of reach, but I'm optimistic.

First a bit more notation/terminology. Let  $\mathcal{G}$  be the group of transformations on  $\mathbb{X}$  described before, still with the simplifying assumption that  $\mathbb{T} = \mathcal{G} = \overline{\mathcal{G}}$ , and let  $\mathcal{K}$  be another group with binary operation  $*$ . Suppose further that  $\mathcal{K}$  is equipped with a total order  $\lesssim$  that's bi-invariant in the sense that, for all  $a, b, k \in \mathcal{K}$ ,

$$a \lesssim b \implies k * a \lesssim k * b \quad \text{and} \quad a * k \lesssim b * k.$$

Let  $\psi : \mathcal{G} \rightarrow \mathcal{K}$  be a homomorphism, so that  $\psi(g \circ g') = \psi(g) * \psi(g')$  for all  $g, g' \in \mathcal{G}$ . Now, for a fixed  $k \in \mathcal{K}$ , define the hypothesis about  $\theta$  as

$$A = \{\theta : \psi(\theta) \lesssim k\}. \quad (13)$$

Then hypotheses of the form (13), determined by  $\psi$  and  $k$ , are not afflicted by false confidence.

**Theorem 5** *For a homomorphism  $\psi$  from  $(\mathcal{G}, \circ)$  to the ordered group  $(\mathcal{K}, *, \lesssim)$ , and for  $k \in \mathcal{K}$ , the hypothesis  $A = A_{\psi, k}$  in (13) is not afflicted with false confidence, i.e.,*

$$\sup_{\theta \in A} \mathbb{P}_\theta\{Q_X(A) \leq \alpha\} \leq \alpha, \quad \text{for all } \alpha \in [0, 1].$$

**Proof** Take  $A$  as in (13) and fix any  $\theta \in A$ . Then  $\psi(\theta) \lesssim k$  and, hence,  $k^{-1} * \psi(\theta) \lesssim e$ , where  $e$  here denotes the identity element in  $\mathcal{K}$ . Recall that the fiducial distribution  $\Theta \sim Q_x$  equals the conditional distribution of  $g \circ H^{-1}$ , given

$U = u$ , which depends on the values  $(g, u)$  of  $t(x)$  but not on  $\theta$ . Since  $\lesssim$  is bi-invariant and  $\psi$  is a homomorphism, we find that  $\psi(\Theta) \lesssim k$  is equivalent to  $\psi(H) \gtrsim k^{-1} * \psi(g)$  and, consequently,

$$\begin{aligned} Q_X(A) \leq \alpha &\iff \text{P}\{\psi(H) \gtrsim k^{-1} * \psi(g) \mid U = y\} \leq \alpha \\ &\iff k^{-1} * \psi(g) \gtrsim z_\alpha(u), \end{aligned}$$

where  $z_\alpha(u)$  is the  $(1 - \alpha)$ -quantile of  $\psi(H)$ , relative to  $\lesssim$ , based on the conditional distribution of  $H$ , given  $U = u$ , derived from  $X \sim P_\theta$ . But the distribution of  $G$ , as a function of  $X \sim P_\theta$ , given that the  $U$ -component of  $t(X)$  is  $u$ , is exactly the distribution of  $\theta \circ H$ , as a function of  $H$ , given  $U = u$ . Therefore,

$$\begin{aligned} \text{P}_\theta\{Q_X(A) \leq \alpha\} &= \text{P}\{k^{-1} * \psi(\theta \circ H) \gtrsim z_\alpha(u) \mid U = u\} \\ &= \text{P}\{k^{-1} * \psi(\theta) * \psi(H) \gtrsim z_\alpha(u) \mid U = u\} \\ &\leq \text{P}\{\psi(H) \gtrsim z_\alpha(u) \mid U = u\}, \end{aligned} \tag{14}$$

where the inequality in (14) follows from the fact that

$$k^{-1} * \psi(\theta) * \psi(H) \gtrsim z_\alpha(u) \iff \psi(H) \gtrsim z_\alpha(u),$$

which, in turn, is a consequence of the fact that

$$\theta \in A \iff \psi(\theta) \lesssim k \iff k^{-1} * \psi(\theta) \lesssim e,$$

where  $e$  is the identity element in  $\mathcal{K}$ . By definition of  $z_\alpha(u)$ , the probability in (14) is no more than  $\alpha$ . ■

I apologize for the level of abstraction, but this formulation helps to pinpoint the kind of structure that's incompatible with false confidence. To make this result more tangible, think of  $(\mathcal{K}, *, \lesssim)$  as real numbers under addition. Then the most natural kinds of homomorphisms in this case would be linear functions of  $\theta$ . Then Theorem 5 says that hypotheses concerning linear functions of  $\theta$  are free from false confidence, e.g., if  $\theta$  is a  $D$ -vector, then the fiducial (and Bayesian) probabilities assigned to hypotheses  $A = \{\theta : \theta_d \leq k\}$  for  $k \in \mathbb{R}$  and  $d = 1, \dots, D$  are reliable. Non-linear functions, such as  $\psi(\theta) = \|\theta\|$  from the application in Balch et al. (2019), may not be homomorphisms and Theorem 5 does not protect the corresponding hypotheses from false confidence. Fraser (2011) also discusses the differences between linear and non-linear functions of  $\theta$  as it relates to reliability of inference.

## 6. Conclusion

For invariant statistical models considered here, it's well-known that the fiducial and default (right Haar) prior Bayes solutions agree and, moreover, that the credible sets derived

from these are exact confidence regions. Under the same setup, I showed here that the fiducial solution can also be viewed as the "best" probabilistic approximation to the valid IM's possibilistic solution. This sheds important new light on the relationship between fiducial and IM solutions. Indeed, if one's only concerned with, say, confidence regions for  $\Theta$ , then both solutions give the same results and, hence, there's no need to bother with the IM's imprecision. But if one is interested in other questions, e.g., inference about certain features  $\psi(\Theta)$  of  $\Theta$ , then the connection between the fiducial and IM solutions is broken and whatever reliability the former might have for inference about  $\Theta$  is lost when marginalized to  $\psi(\Theta)$ .

Beyond invariant statistical models, one could *define* a fiducial distribution to be the maximal probabilistic approximation to the IM solution's possibilistic output. Just like in the invariant model case, this definition of a fiducial distribution would ensure that fiducial credible regions are exact confidence regions. But this strong calibration comes at a computational cost. In that case, one might prefer an existing constructive solution, e.g., generalized fiducial, but this would likely not be a member of the IM's credal set and, therefore, the aforementioned confidence-connection can at best be achieved in an asymptotically approximate sense. And just like above, the strong validity property achieved by the IM is out of reach for all fiducial solutions.

Since the validity property is fundamental to reliable uncertainty quantification, my claim is that the IM solution is objectively better than the Bayes/fiducial solution. So, if one opts for the latter instead of the former, then it's solely for the familiarity and/or convenience of ordinary probability theory, and this doesn't come without the cost of a potential loss of reliability. As the computational tools for evaluating the IM output continue to develop, this convenience gap will likewise continue to close, to the point that there's no justification to sacrifice on reliability.

## Acknowledgments

This work is supported by the U.S. National Science Foundation, grant SES-2051225.

## References

- Thomas Augustin, Gero Walter, and Frank P. A. Coolen. Statistical inference. In *Introduction to Imprecise Probabilities*, Wiley Ser. Probab. Stat., pages 135–189. Wiley, Chichester, 2014.
- Michael S. Balch, Ryan Martin, and Scott Ferson. Satellite conjunction analysis and the false confidence theorem. *Proc. Royal Soc. A*, 475(2227):2018.0565, 2019.

- James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, second edition, 1985.
- James O. Berger, José M. Bernardo, and Dongchu Sun. The formal definition of reference priors. *Ann. Statist.*, 37(2): 905–938, 2009.
- Jose-M. Bernardo. Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. Ser. B*, 41:113–147, 1979.
- 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. ISBN 0-471-92416-4.
- Inés Couso, Susana Montes, and Pedro Gil. The necessity of the strong  $\alpha$ -cuts of a fuzzy set. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, 9(2):249–262, 2001.
- Y. Cui and J. Hannig. Demystifying inferential moels: A fiducial approach. [arXiv:2205.05612](https://arxiv.org/abs/2205.05612), 2022.
- Céline Cunen, Nils Lid Hjort, and Tore Schweder. Confidence in confidence distributions! *Proc. Roy. Soc. A*, 476:20190781, 2020.
- A. P. Dawid. Fiducial inference then and now. [arXiv:2012.10689](https://arxiv.org/abs/2012.10689), 2020.
- A. P. Dawid and M. Stone. The functional-model basis of fiducial inference. *Ann. Statist.*, 10(4):1054–1074, 1982. ISSN 0090-5364. With discussion.
- A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38:325–339, 1967.
- A. P. Dempster. A generalization of Bayesian inference. (With discussion). *J. Roy. Statist. Soc. Ser. B*, 30:205–247, 1968. ISSN 0035-9246.
- A. P. Dempster. The Dempster–Shafer calculus for statisticians. *Internat. J. Approx. Reason.*, 48(2):365–377, 2008.
- Sébastien Destercke and Didier Dubois. Special cases. In *Introduction to Imprecise Probabilities*, Wiley Ser. Probab. Stat., pages 79–92. Wiley, Chichester, 2014.
- Didier Dubois, Laurent Foulloy, Gilles Mauris, and Henri Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliab. Comput.*, 10(4):273–297, 2004.
- Morris L. Eaton. *Group Invariance Applications in Statistics*. Institute of Mathematical Statistics, Hayward, CA, 1989. ISBN 0-940600-15-3.
- Bradley Efron. R. A. Fisher in the 21st century. *Statist. Sci.*, 13(2):95–122, 1998. ISSN 0883-4237.
- Bradley Efron. Discussion: “Confidence distribution, the frequentist distribution estimator of a parameter: a review” [mr3047496]. *Int. Stat. Rev.*, 81(1):41–42, 2013.
- R. A. Fisher. Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 26:528–535, 1930.
- R. A. Fisher. The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proc. R. Soc. Lond. A.*, 139:343–348, 1933.
- R. A. Fisher. The fiducial argument in statistical inference. *Ann. Eugenics*, 6:391–398, 1935.
- D. A. S. Fraser. *The Structure of Inference*. John Wiley & Sons Inc., New York, 1968.
- D. A. S. Fraser. Is Bayes posterior just quick and dirty confidence? *Statist. Sci.*, 26(3):299–316, 2011.
- D. A. S. Fraser, M. Bédard, A. Wong, Wei Lin, and A. M. Fraser. Bayes, reproducibility and the quest for truth. *Statist. Sci.*, 31(4):578–590, 2016.
- Jayanta K. Ghosh, Mohan Delampady, and Tapas Samanta. *An Introduction to Bayesian Analysis*. Springer, New York, 2006.
- Paul R. Halmos. *Measure Theory*. D. Van Nostrand Co., Inc., New York, N. Y., 1950.
- Jan Hannig, Hari Iyer, Randy C. S. Lai, and Thomas C. M. Lee. Generalized fiducial inference: a review and new results. *J. Amer. Statist. Assoc.*, 111(515):1346–1361, 2016.
- Dominik Hose, Michael Hanss, and Ryan Martin. A practical strategy for valid partial prior-dependent possibilistic inference. In S. Le Hégarat-Masclé, I. Bloch, and E. Aldea, editors, *Belief Functions: Theory and Applications (BELIEF 2022)*, volume 13506 of *Lecture Notes in Artificial Intelligence*, pages 197–206. Springer, 2022.
- Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London Ser. A*, 186:453–461, 1946. ISSN 0962-8444.
- Yi Lin, Ryan Martin, and Min Yang. On optimal designs for nonregular models. *Ann. Statist.*, 47(6):3335–3359, 2019. ISSN 0090-5364.

- Chuanhai Liu and Ryan Martin. Inferential models and possibility measures. *Handbook of Bayesian, Fiducial, and Frequentist Inference*, to appear; arXiv:2008.06874, 2021.
- Kanti V. Mardia and Peter E. Jupp. *Directional Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2000. ISBN 0-471-95333-4.
- Ryan Martin. False confidence, non-additive beliefs, and valid statistical inference. *Internat. J. Approx. Reason.*, 113:39–73, 2019.
- Ryan Martin. An imprecise-probabilistic characterization of frequentist statistical inference. arXiv:2112.10904, 2021a.
- Ryan Martin. Inferential models and the decision-theoretic implications of the validity property. arXiv:2112.13247, 2021b.
- Ryan Martin. Valid and efficient imprecise-probabilistic inference under partial priors, I. First results. arXiv:2203.06703, 2022a.
- Ryan Martin. Valid and efficient imprecise-probabilistic inference under partial priors, II. General framework. arXiv:2211.14567, 2022b.
- Ryan Martin and Chuanhai Liu. Inferential models: a framework for prior-free posterior probabilistic inference. *J. Amer. Statist. Assoc.*, 108(501):301–313, 2013.
- Ryan Martin and Chuanhai Liu. *Inferential Models*, volume 147 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2015.
- Ryan Martin, Michael Balch, and Scott Ferson. Response to the comment ‘Confidence in confidence distributions!’. *Proc. R. Soc. A.*, 477:20200579, 2021.
- Alexander Murph, Jan Hannig, and Jonathan P. Williams. Introduction to generalized fiducial inference. *Handbook of Bayesian, Fiducial, and Frequentist Inference*, to appear., 2021.
- Leopoldo Nachbin. *The Haar Integral*. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London, 1965.
- Saralees Nadarajah, Sergey Bityukov, and Nikolai Krasnikov. Confidence distributions: a review. *Stat. Methodol.*, 22: 23–46, 2015. ISSN 1572-3127.
- J. Neyman. Fiducial argument and the theory of confidence intervals. *Biometrika*, 32:128–150, 1941.
- Mark J. Schervish. *Theory of Statistics*. Springer-Verlag, New York, 1995. ISBN 0-387-94546-6.
- Teddy Seidenfeld. R. A. Fisher’s fiducial argument and Bayes’ theorem. *Statist. Sci.*, 7(3):358–368, 1992. ISSN 0883-4237.
- Arkady Shemyakin. Hellinger distance and non-informative priors. *Bayesian Anal.*, 9(4):923–938, 2014. ISSN 1936-0975.
- Nicholas Syring and Ryan Martin. Stochastic optimization for numerical evaluation of imprecise probabilities. In Andrés Cano, Jasper De Bock, Enrique Miranda, and Serafín Moral, editors, *Proceedings of the Twelfth International Symposium on Imprecise Probability: Theories and Applications*, volume 147 of *Proceedings of Machine Learning Research*, pages 289–298. PMLR, 2021.
- Gunnar Taraldsen. Joint confidence distributions. <https://doi.org/10.13140/RG.2.2.33079.85920>, 2021.
- Gunnar Taraldsen and Bo Henry Lindqvist. Fiducial theory and optimal inference. *Ann. Statist.*, 41(1):323–341, 2013. ISSN 0090-5364.
- Suzanne Thornton and Min-ge Xie. Bridging Bayesian, frequentist and fiducial (BFF) inferences using confidence distribution. arXiv:2012.04464, 2020.
- A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Cambridge, 1998. ISBN 0-521-49603-9; 0-521-78450-6.
- Peter Walley. *Statistical Reasoning with Imprecise Probabilities*, volume 42 of *Monographs on Statistics and Applied Probability*. Chapman & Hall Ltd., London, 1991.
- Minge Xie and Kesar Singh. Confidence distribution, the frequentist distribution estimator of a parameter: a review. *Int. Stat. Rev.*, 81(1):3–39, 2013.
- S. L. Zabell. R. A. Fisher and the fiducial argument. *Statist. Sci.*, 7(3):369–387, 1992.