Which statistical hypotheses are afflicted with false confidence?

Ryan Martin*

North Carolina State University
Department of Statistics
Raleigh, North Carolina 27695, U.S.A.
rgmarti3@ncsu.edu

Abstract. The false confidence theorem establishes that, for any data-driven, precise-probabilistic method for uncertainty quantification, there exists (both trivial and non-trivial) false hypotheses to which the method tends to assign high confidence. This raises concerns about the reliability of these widely-used methods, and shines promising light on the consonant belief function-based methods that are provably immune to false confidence. But an existence result alone leaves much to be desired. Towards an answer to the title question, I show that, roughly, complements of convex hypotheses are afflicted by false confidence.

Keywords: Bayesian; consonant beliefs; convexity; inferential model; fiducial inference; possibility theory; validity.

1 Introduction

In *Logic of Statistical Inference*, Hacking (1976) writes: "Statisticians want numerical measures of the degree to which data support hypotheses." One such measure is a Bayesian posterior probability, but imprecise probabilists—the belief function community specifically—are well aware that precise probability theory is not the only mode of uncertainty quantification. Indeed, in a statistical inference problem, where prior information is at best incomplete and data speaks only indirectly through a model, there's good reason to question the appropriateness and/or reliability of a precise probability as statisticians' go-to quantitative expression of the degree to which data supports hypotheses.

Balch et al. (2019) expressed this concern in terms of *false confidence*. Roughly, in the context described in Section 2.1, false confidence corresponds to the existence of false hypotheses to which, say, a default-prior Bayesian posterior distribution tends to assign high probability, support, or confidence. Their result also applies to (generalized) fiducial inference (Dawid 2020; Fisher 1935; Hannig et al. 2016), confidence distributions (Xie and Singh 2013), etc., so it highlights a risk of unreliability inherent in *all* precise-probabilistic approaches to statistical uncertainty quantification. Since reliability is obviously a top priority, there's

* Partially supported by the U.S. National Science Foundation, SES–205122.

an exciting opportunity for imprecise probability theory to make a fundamental contribution to statistics, a domain in which imprecise-probabilistic methods are greatly under-appreciated and largely unused. Along these lines, I've recently shown (Martin 2019, 2021, 2022a,b) that a suitable *possibilistic*, or *consonant belief* framework for statistical inference is immune to false confidence; that is, this framework is reliable in the sense that it provably doesn't tend to assign high support to any false hypotheses!

Unfortunately, the *false confidence theorem*, as stated in Balch et al. (2019), is only an existence result. In a certain sense, the existence of hypotheses that are afflicted with false confidence is "obvious," and it's partly for this reason that statisticians largely haven't taken this too seriously (Carmichael and Williams 2018; Cunen et al. 2020; Martin et al. 2021). But the extent of false confidence affliction goes well beyond the hypotheses for which it's obvious: this has been demonstrated empirically in a number of specific examples, but no theoretical characterizations have been put forward. To my knowledge, all that's currently known is, for location-scale and other group transformation models, linear hypotheses about the uncertain Θ of the form "$a^\top\Theta \leq b$" are safe from false confidence (Martin 2023a). So, theoretically, we currently know effectively nothing about which hypotheses are afflicted with false confidence, but the present makes some progress in this direction. In particular, under a simple model that (approximately) represents most practical cases, I show that a class of (non-linear) hypotheses which includes those that are *co-convex*, i.e., complements of convex sets, are afflicted with false confidence. This is not a complete characterization, but it provides some insight as to what structure breeds false confidence.

Admittedly, the present paper says very little about belief functions and imprecise probability, but I still expect this investigation to make a significant contribution. Indeed, once the extent and implications of false confidence are understood, statisticians who care about reliable uncertainty quantification will have no choice but to use certain imprecise-probabilistic ideas and methods.

2 Background

2.1 Problem setup

Let X denote the data taking values in a general sample space $\mathbb{X}$. A statistical model consists of a family of probability distributions $\{\mathsf{P}_\theta : \theta \in \mathbb{T}\}$ on $\mathbb{X}$ indexed by a general parameter space $\mathbb{T}$. As is commonly assumed, suppose there is an uncertain "true" parameter value Θ such that X has distribution P_Θ. I'll assume throughout that prior information about Θ is vacuous. The high-level goal is to quantify uncertainty about Θ, given $X = x$, à la Hacking.

2.2 Inferential models

Following Martin (2021), an inferential model (IM) is a map from data $x \in \mathbb{X}$ to a lower probability $\underline{\Pi}_x$ supported on subsets of $\mathbb{T}$, which depends implicitly on the statistical model and perhaps other things, e.g., prior information

about Θ. The interpretation is that $\underline{\Pi}_x(H)$ measures the degree of support for or belief/confidence in the truthfulness of the hypothesis "$\Theta \in H$" given data $X = x$. Examples of precise IMs include Bayes's posterior probabilities, Fisher's fiducial distributions, and Hannig's generalized fiducial distributions; examples of imprecise IMs include Dempster's seminal proposal (Dempster 1966, 1968, 2008), Walley's generalized Bayes (Walley 1991), consonant likelihood-based belief functions (Denœux 2014; Shafer 1982; Wasserman 1990), and what's briefly described in Section 2.4 below.

Bayesian- and fiducial-like frameworks quantify uncertainty about Θ, given $X = x$, with a precise (countably additive) "posterior distribution"

$$\Pi_x(H) = \frac{\int_H L_x(\theta)\,\Pi(d\theta)}{\int_{\mathbb{T}} L_x(\theta)\,\Pi(d\theta)}, \quad H \subseteq \mathbb{T}, \tag{1}$$

where Π is like a "prior distribution" for the uncertain parameter Θ and $\theta \mapsto L_x(\theta)$ is the model's likelihood function given $X = x$. The quotation marks are intended to highlight the point that, since prior information is assumed vacuous, these are "prior" and "posterior" distributions only in a formal sense. As Jeffreys (1946) explains, Π is a default measure that gets updated to Π_x by formally following Bayes's rule when $X = x$. In certain contexts (e.g., Hannig et al. 2016), Π itself might depend on data, hence can't represent genuine prior information. In any case, the map $H \mapsto \Pi_x(H)$ is often used in applications to quantify uncertainty about Θ, given $X = x$. But "[Bayes's rule] does not create real probabilities from hypothetical probabilities" (Fraser 2014), so a practically and theoretically important question is if this brand of (precise) probabilistic uncertainty quantification is reliable.

2.3 False confidence

The false confidence theorem (Balch et al. 2019; Martin 2019) says that, for *any precise IM*, i.e., a mapping $x \mapsto \Pi_x$, with Π_x a probability measure on $\mathbb{T}$, there exists a hypothesis–threshold pair (H, α) such that

$$H \not\ni \Theta \quad \text{and} \quad \mathsf{P}_\Theta\{\Pi_X(H) \geq 1 - \alpha\} > \alpha. \tag{2}$$

That is, there exists false hypotheses H to which the posterior tends to assign a relatively large probability/confidence, shedding light on a lurking unreliability. That is, the statistician would tend to be confident in a hypothesis based on data X if its Π_X-probability is relatively large, but this is unreliable if $\Pi_X(H)$ tends to be relatively large even if H is false.

Martin (2023b) shows that false confidence implies an incoherence-like risk of monetary loss to statisticians who quantify uncertainty using precise IMs. To see this, consider the following class of (contingent) gambles

$$f_\theta^{H,\alpha}(x) = \begin{cases} 1(\theta \in H) - (1 - \alpha) & \text{if } \Pi_x(H) > 1 - \alpha \\ 0 & \text{otherwise,} \end{cases}$$

where $1(\cdot)$ is the indicator function. For every (α, H, x, θ), this gamble would be acceptable to the statistician who quantifies his uncertainty with Π_x when he observes $X = x$; that is, the expected value of $f_\Theta^{H,\alpha}(x)$, with respect to Π_x for any fixed x, is positive. Now imagine another agent, a scrutinizer, who doubts the reliability of the statistician's claims. False confidence creates an opportunity for this scrutinizer—through careful considerations, background knowledge, or simply luck—to force the statistician into a systematic loss. If (H, α) is one of the hypothesis–threshold pairs that satisfies (2), then, as a function of X for fixed $\Theta \notin H$, the statistician's winnings $f_\Theta^{H,\alpha}(X)$ are either negative (with probability α) or zero. Therefore, his "long-run" earnings are negative, hence a systematic loss. Note that "long-run" doesn't require replications of a given experiment under the same settings or even by the same statistician. If groups of statisticians quantify their uncertainty using a precise IM, then scrutinizers can, in principle, make the statisticians collectively systematic losers. Also, the scrutinzers don't need to *know* the unknown Θ to force this systematic loss, they only need to find, even just by luck, hypotheses afflicted by false confidence. If, as I claim, hypotheses afflicted by false confidence aren't uncommon, then the above points ought to raise concern.

2.4 Consonant beliefs to the rescue

Fisher (1930) writes: "...[the likelihood function] does not obey the laws of probability; it involves no differential element." The default prior Π also has no meaningful differential element "$d\theta$"—with vacuous prior information, there's no reason to think that measure-theoretically larger hypotheses are "more likely" than smaller ones. However, if neither the likelihood nor the prior have a meaningful differential element, then there's no sense in which the differential element on the right-hand side of (1) could be meaningful. Indeed, it's easy to find (measure-theoretically) large hypotheses that are false, hence the trivial cases of false confidence. More generally, I claim that the meaningless differential element is at the heart of false confidence (Martin 2023c).

If there exists an IM that protects all hypotheses from false confidence, then it must be imprecise; the remarks in the previous paragraph suggest that it should also be differential element free. The simplest example of this is a consonant belief function (Shafer 1976), one whose conjugate plausibility function is a maxitive possibility measure (Dubois and Prade 1988; Hose 2022). To my knowledge, the first IMs shown to be *valid*, i.e.,

$$\sup_{\Theta \notin H} \mathsf{P}_\Theta\{\underline{\Pi}_X(H) \geq 1-\alpha\} \leq \alpha, \quad \text{for all } (H, \alpha), \tag{3}$$

were those put forward in Martin and Liu (2015) based on nested random sets; see, also, Balch (2012) and Denœux and Li (2018). The condition (3) implies, among other things, that there's no false confidence. The valid IM construction has been generalized and streamlined in Martin (2015, 2018, 2022b), but those specific details won't be needed in what follows.

3 Co-convexity breeds false confidence

Without much loss of generality, I'll focus here on the D-dimensional Gaussian case $X \sim \mathsf{N}_D(\Theta, \Sigma)$, where $\Theta \in \mathbb{T} = \mathbb{R}^D$ is the uncertain parameter and the $D \times D$ covariance matrix Σ is fixed and known. Then the likelihood is

$$L_X(\theta) \propto \exp\{-\tfrac{1}{2}(X-\theta)^\top \Sigma^{-1}(X-\theta)\}, \quad \theta \in \mathbb{T}.$$

I say "without much loss of generality" because, in most of the statistical models used in practical applications, there's a corresponding Gaussian limit experiment (e.g., van der Vaart 1998, Chapter 9). That is, if the sample size is large, then the maximum likelihood estimator (say) is an approximately minimal sufficient statistic whose sampling distribution is approximately Gaussian with mean Θ and covariance matrix a multiple of the inverse Fisher information. In this case, with vacuous prior information, the go-to precise IM for Θ is

$$\Pi_X = \mathsf{N}_D(X, \Sigma). \tag{4}$$

The precise IM in (4) has a number of desirable properties, e.g, highest posterior density credible sets are minimum volume confidence sets. But it still suffers from the inherent unreliability exposed by the false confidence theorem.

To develop some intuition, consider a function $\phi : \mathbb{T} \to \mathbb{R}$, and define

$$H_\phi = \{\theta \in \mathbb{T} : \phi(\theta) > \phi(\Theta)\}. \tag{5}$$

Clearly, hypothesis H_ϕ is *false*, i.e., $H_\phi \not\ni \Theta$. If ϕ is (non-linear) convex, which makes H_ϕ *co-convex*—the complement of a convex set—then Jensen's inequality immediately gives the bound $\mathsf{E}_\Theta\{\phi(X)\} > \phi(\Theta)$. Consequently, there must be non-negligible probability that X, the Π_X-posterior mean, is contained in the false H_ϕ; and, if the posterior mean is in H_ϕ, then the corresponding posterior probability, $\Pi_X(H_\phi)$, can't be small, hence an ample opportunity for false confidence in H_ϕ. Interestingly, this apparently has little to do with the size/measure of H_ϕ or of H_ϕ^c: something else is driving false confidence.

The more general, albeit less intuitive, result is presented next. I'll start with a definition. A set $G \subset \mathbb{T}$ will be called *non-linear, locally convex at* ϑ, or ϑ-*noloco*, if it satisfies the following three properties:

- if G contains ϑ on its boundary,
- if it has a supporting hyperplane at ϑ, and
- if the intersection of G^c with the half-space determined by the supporting hyperplane that contains G has non-zero Lebesgue measure.

To connect this to the more intuitive discussion above, if ϕ is a non-linear convex function, then the complement of H_ϕ in (5) is Θ-noloco. More generally, if G is convex, then a supporting hyperplane exists at each of its boundary points (e.g., Boyd and Vandenberghe 2004, Sec. 2.5.2). But G could be non-convex and have a supporting hyperplane at some of its boundary points—Figure 1 shows a

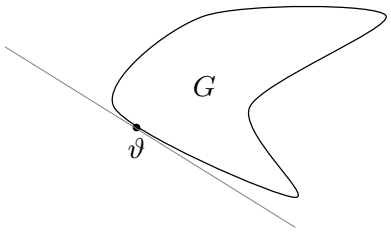


Fig. 1. A non-convex G that's ϑ-noloco; gray line defines the supporting hyperplane.

non-convex G that's still ϑ-noloco. The aforementioned supporting hyperplane at ϑ is determined by a vector g_ϑ, i.e., that hyperplane is

$$\{\theta \in \mathbb{T} : g_\vartheta^\top (\theta - \vartheta) = 0\},$$

and the half-space that contains G is

$$\text{halfsp}_\vartheta(G) = \{\theta \in \mathbb{T} : g_\vartheta^\top (\theta - \vartheta) \leq 0\}.$$

If the boundary of G was linear, then its boundary would coincide with the boundary of the half-space defined above. The analysis below requires that there is some room (having non-zero Lebesgue measure) between the boundary of G and the boundary of the half-space, which is enforced by the third condition above. This non-linearity and boundary separation is shown in Figure 1.

Proposition 1. *For any $\Theta \in \mathbb{T}$, if G is Θ-noloco, then the hypothesis $H = G^c$ is afflicted by false confidence. In particular, the random variable $\Pi_X(H)$, as a function of $X \sim \mathsf{N}_D(\Theta, \Sigma)$, is stochastically larger than $\mathsf{Unif}(0,1)$.*

Proof. Let g_Θ denote the vector that defines the supporting hyperplane of G at Θ. Since G is contained in the half-space $\text{halfsp}_\Theta(G)$, we get

$$H \supset H_{\text{lin}} := \{\theta \in \mathbb{T} : g_\Theta^\top (\theta - \Theta) > 0\},$$

and, consequently, $\Pi_X(H) > \Pi_X(H_{\text{lin}})$. The last inequality is strict because Π_X is absolutely continuous with respect to Lebesgue measure and, by assumption, $H \setminus H_{\text{lin}}$ has positive Lebesgue measure. The lower bound, $\Pi_X(H_{\text{lin}})$, satisfies

$$\Pi_X(H_{\text{lin}}) = 1 - F\Big(-\frac{g_\Theta^\top (X - \Theta)}{\{g_\Theta^\top \Sigma\, g_\Theta\}^{1/2}}\Big), \tag{6}$$

where F is the standard normal distribution function. As a function of $X \sim \mathsf{N}_D(\Theta, \Sigma)$, the right-hand side of (6) is $\mathsf{Unif}(0,1)$. Therefore, $\Pi_X(H)$ is (strictly) lower-bounded by a $\mathsf{Unif}(0,1)$ random variable, completing the proof.

By no means is this a complete characterization of false confidence. For one thing, it's absolutely not necessary for Θ to sit on the boundary of the

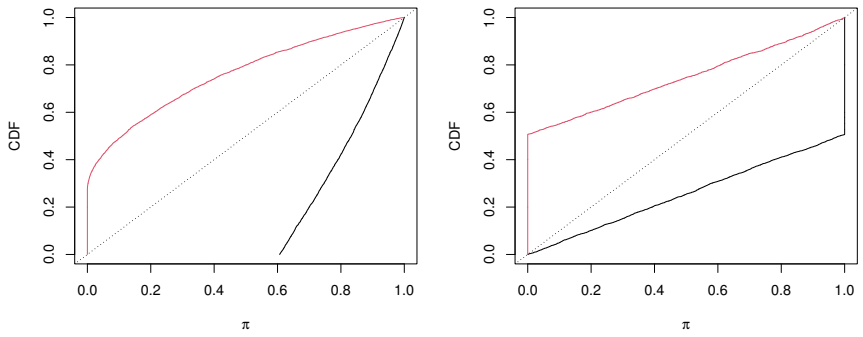
hypothesis—I imposed this constraint just to make the analysis tractable. Similar results are expected for hypotheses that miss Θ but not by too much. More generally, I don't believe that noloco is fundamental to false confidence. My conjecture is that all *non-linear* hypotheses about Θ, e.g., "$\phi(\Theta) \leq b$" for a non-linear map ϕ, have at least a mild case of false confidence—the reason being that non-linear mapping can warp the parameter space in such a way that probability assignments get pushed in one direction or another systematically. Precisely diagnosing the existence and severity of affliction remains an open question.

4 Illustrations

In this section, I present two examples showing the existence and severity of false confidence. Both illustrations are rather simple, but they're still forceful. Indeed, if the manifestation of false confidence is relatively easy to spot in these simple examples, then we can be sure that it's present in complex, modern applications too. It's for precisely this reason that the statistical community shouldn't ignore these warnings, assume that false confidence is too rare to be concerned about, and stick with the (Bayesian) status quo.

Example 1. Non-linear hypotheses. Inference on the squared length of a normal mean vector is a classically challenging statistical problem, originating in Stein (1959) and appearing as the late D. R. Cox's *Challenge Question E* (Fraser et al. 2018). It's also closely related to the motivating satellite collision example in Balch et al. (2019). In the present context, if $\phi(\theta) = \|\theta\|^2$, then the set H_ϕ defined in (5) determines a (false) hypothesis about the mean vector's squared length; this set is also co-convex and, therefore, by Proposition 1, is afflicted with false confidence. To see the extent of affliction, the CDF $\pi \mapsto \mathsf{P}_\Theta\{\Pi_X(H_\phi) \leq \pi\}$ is shown in Figure 2(a), where the dimension is $D = 2$ and Θ is length 1. Note that, in this case, $\Pi_X(H_\phi)$ is always greater than 0.6, even though H_ϕ is false. For comparison, Figure 2(a) also displays the CDF of the valid IM's lower probability $\underline{\Pi}_X(H_\phi)$ and, since it lies above the diagonal line corresponding to the $\mathsf{Unif}(0,1)$ CDF, there's clearly no false confidence.

Example 2. Non-linear parameter space. Fraser (2011) considers a normal mean model $X \sim \mathsf{N}(\Theta, 1)$ but with the side information that Θ has a *known* lower bound, which I take to be 0 without loss of generality. This is motivated by relevant high-energy particle physics applications, but I'll simply point the reader to, e.g., Mandelkern (2002) for these details. Consider the (false) hypothesis $H = (\Theta, \infty)$—but note that the parameter constraint makes H^c bounded. What's interesting about this example is that, apparently, the bounded and, hence, non-linear parameter space forces non-linearity in an otherwise linear hypothesis, which induces false confidence. So, this example offers a glimpse into the breadth and diversity of cases where false confidence can emerge, perhaps unexpectedly. The CDFs of the (flat-prior) Bayesian posterior probability, $\Pi_X(H)$, and of the valid IM's lower probability, $\underline{\Pi}_X(H)$, are shown in Figure 2(b), based on true



(a) Example 1 (b) Example 2

Fig. 2. Black lines are CDFs for the (flat-prior) Bayes posterior probabilities and red lines are the CDFs for the corresponding valid IM's lower probabilities.

$\Theta = 1$. Note that the Bayes posterior assigns probability 1 to the false hypothesis 50% of the time, while the valid IM does the polar opposite, rightfully assigning 0 (or small) support to the false hypothesis most of the time.

5 Conclusion

The present paper is concerned with the following question: *which statistical hypotheses are afflicted by false confidence?* My collaborators and I have had intuition about how to answer this question for some time, but only now have I been able to formulate this intuition in a way that's conducive to mathematical analysis. The result that I proved here is quite simple, perhaps unremarkable, but I'd argue that simplicity is a virtue. After all, false confidence is the rule, rather than the exception, so it should be easy to identify hypotheses that are afflicted. What's interesting is that a property slightly more general than co-convexity is what makes the hypothesis vulnerable to false confidence.

The result presented here provides a sufficient condition for false confidence, but I seriously doubt that the same condition is necessary. As above, my conjecture is that non-linearity is enough to create at least a susceptibility to false confidence. Non-linearlity alone may not be severe enough to cause false confidence-level problems as defined here; but maybe to cause the milder but still concerning "fluke confidence" that my friend and collaborator, Michael Balch, has been telling me about recently. In any case, the advancements made in the present paper make me optimistic that we'll soon be able to settle these questions.

Acknowledgments. Thanks to the reviewers for their helpful comments. The author is partially supported by the U.S. National Science Foundation, SES–2051225.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

Bibliography

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

Balch, M. S., Martin, R., and Ferson, S. (2019). Satellite conjunction analysis and the false confidence theorem. *Proc. Royal Soc. A*, 475(2227):2018.0565.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge.

Carmichael, I. and Williams, J. P. (2018). An exposition of the false confidence theorem. *Stat*, 7(1):e201.

Cunen, C., Hjort, N. L., and Schweder, T. (2020). Confidence in confidence distributions! *Proc. Roy. Soc. A*, 476:20190781.

Dawid, A. P. (2020). Fiducial inference then and now. `arXiv:2012.10689`.

Dempster, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *Ann. Math. Statist.*, 37:355–374.

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

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

Denœux, T. (2014). Likelihood-based belief function: justification and some extensions to low-quality data. *Internat. J. Approx. Reason.*, 55(7):1535–1547.

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

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

Fisher, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 26:528–535.

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

Fraser, D. A. S. (2011). Is Bayes posterior just quick and dirty confidence? *Statist. Sci.*, 26(3):299–316.

Fraser, D. A. S. (2014). Why does statistics have two theories? In Lin, X., Genest, C., Banks, D. L., Molenberghs, G., Scott, D. W., and Wang, J.-L., editors, *Past, Present, and Future of Statistical Science*, chapter 22. Chapman & Hall/CRC Press.

Fraser, D. A. S., Reid, N., and Lin, W. (2018). When should modes of inference disagree? Some simple but challenging examples. *Ann. Appl. Stat.*, 12(2):750–770.

Hacking, I. (1976). *Logic of Statistical Inference*. Cambridge University Press, Cambridge-New York-Melbourne.

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

Hose, D. (2022). *Possibilistic Reasoning with Imprecise Probabilities: Statistical Inference and Dynamic Filtering*. PhD thesis, University of Stuttgart.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London Ser. A*, 186:453–461.

Mandelkern, M. (2002). Setting confidence intervals for bounded parameters. *Statist. Sci.*, 17(2):149–172. With comments.

Martin, R. (2015). Plausibility functions and exact frequentist inference. *J. Amer. Statist. Assoc.*, 110(512):1552–1561.

Martin, R. (2018). On an inferential model construction using generalized associations. *J. Statist. Plann. Inference*, 195:105–115.

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

Martin, R. (2021). An imprecise-probabilistic characterization of frequentist statistical inference. `arXiv:2112.10904`.

Martin, R. (2022a). Valid and efficient imprecise-probabilistic inference with partial priors, I. First results. `arXiv:2203.06703`.

Martin, R. (2022b). Valid and efficient imprecise-probabilistic inference with partial priors, II. General framework. `arXiv:2211.14567`.

Martin, R. (2023a). Fiducial inference viewed through a possibility-theoretic inferential model lens. In Miranda, E., Montes, I., Quaeghebeur, E., and Vantaggi, B., editors, *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*, volume 215 of *Proceedings of Machine Learning Research*, pages 299–310. PMLR.

Martin, R. (2023b). Fisher's underworld and the behavioral–statistical reliability balance in scientific inference. `arXiv:2312.14912`.

Martin, R. (2023c). A possibility-theoretic solution to Basu's Bayesian–frequentist via media. *Sankhya A*, to appear, `arXiv:2303.17425`.

Martin, R., Balch, M., and Ferson, S. (2021). Response to the comment 'Confidence in confidence distributions!'. *Proc. R. Soc. A.*, 477:20200579.

Martin, R. and Liu, C. (2015). *Inferential Models*, volume 147 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N.J.

Shafer, G. (1982). Belief functions and parametric models. *J. Roy. Statist. Soc. Ser. B*, 44(3):322–352. With discussion.

Stein, C. (1959). An example of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Statist.*, 30:877–880.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

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

Wasserman, L. A. (1990). Belief functions and statistical inference. *Canad. J. Statist.*, 18(3):183–196.

Xie, M. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: a review. *Int. Stat. Rev.*, 81(1):3–39.