

Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: www.tandfonline.com/journals/tstf20

Inferential models: the power of auxiliary variables for reasoning with scientific uncertainty

Chuanhai Liu

To cite this article: Chuanhai Liu (20 Nov 2026): Inferential models: the power of auxiliary variables for reasoning with scientific uncertainty, Statistical Theory and Related Fields, DOI: 10.1080/24754269.2026.2765040

To link to this article: https://doi.org/10.1080/24754269.2026.2765040

© 2026 Purdue University. Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 20 Nov 2026.

Submit your article to this journal

Article views: 30

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tstf20

STATISTICAL THEORY AND RELATED FIELDS
https://doi.org/10.1080/24754269.2026.2765040

Taylor & Francis
by informa

OPEN ACCESS Check for updates

Inferential models: the power of auxiliary variables for reasoning with scientific uncertainty

Chuanhai Liu

Department of Statistics, Purdue University, West Lafayette, USA

ABSTRACT

A central challenge in scientific inference is to produce uncertainty assessments that are both situation-specific and frequency-calibrated. This article examines inferential models (IMs) as a framework for prior-free probabilistic reasoning with scientific uncertainty. The central IM idea is to view the auxiliary variables in a sampling model as the source of model-based uncertainty. R. A. Fisher's fiducial inference transfers auxiliary randomness to the parameter space before applying probability calculus; IMs instead predict the unobserved auxiliary value with calibrated predictive random sets (PRSs) and transfer the resulting plausibility statements only afterward. This change in order yields valid uncertainty assessments and clarifies the relations among Fisherian fiducial reasoning, Neymanian confidence theory, Dempster-Shafer belief functions, generalized fiducial inference, and IMs. By comparing IMs with objective-prior Bayesian inference, the article argues that E. T. Jaynes' logic-of-science ambition can be continued without forcing all scientific uncertainty into a precise prior distribution because calibrated imprecision is often essential. Finally, the article suggests that a differential-geometric theory of IMs may be within reach, offering a possible route to foundational questions traditionally framed in terms of the likelihood principle.

ARTICLE HISTORY
Received 18 June 2026
Accepted 12 November 2026

KEYWORDS
Confidence intervals; Dempster-Shafer theory; fiducial inference; Jeffreys' prior; objective priors

1. Introduction

Across its philosophical, methodological, and practical dimensions, as a discipline concerned with transforming empirical experience into knowledge and guiding its application, statistics has long been pulled between two demands. Scientific inference must be conditional on the data actually observed: after seeing the data, scientists seek to learn about their questions of interest from the evidence at hand. At the same time, scientific uncertainty quantification must be frequency-calibrated. A procedure that returns data-driven numerical uncertainty statements that are systematically overconfident or lacking a standardized scale for interpretation cannot be reliably used for the creation of scientific evidence. Bayesian inference satisfies the first demand elegantly when genuine prior information is available. Frequentist inference, in practice, satisfies the second by calibrating procedures through repeated-sampling properties. But these two successes reveal complementary limitations. Bayesian

CONTACT Chuanhai Liu chuanhai@purdue.edu Department of Statistics, Purdue University, West Lafayette, IN 47907, USA

© 2026 Purdue University. Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

inference requires a joint prior distribution over all unknown quantities, whereas frequentist guarantees are often attached to procedures rather than to direct post-data measures of evidence.

The treatment of Bayes in this review follows from that tension. The aim is not to dismiss Bayesian probability, or Jaynes' broad view in *Probability Theory: The Logic of Science* that probability extends logic to uncertain reasoning (Jaynes, 2003). It is to separate the formal logic of probability from the scientific interpretation of a particular prior. Information-based priors, such as Jeffreys' invariant prior, Jaynes' maximum entropy priors, and Bernardo's reference priors, are valuable mathematical devices. But when no genuine prior information is available, their posterior probabilities should not automatically be read as calibrated scientific uncertainty. One goal of this article is to give readers enough inferential model (IM) background to understand both the power and the limits of such information-based Bayesian prior constructions.

R. A. Fisher's fiducial argument enters at precisely this point (Fisher, 1935, 1973). Fisher rejected Bayesian posteriors with arbitrary priors, but he still wanted post-data uncertainty statements about fixed unknowns. Neyman's confidence theory supplied mathematical clarity by placing probability on the random interval procedure, not on probabilistic statements about the fixed parameter (Neyman, 1941). Dempster-Shafer theory introduced lower and upper probabilities generated by modeling on the power space of our usual probability space (Dempster, 2008; Shafer, 1976). Generalized fiducial inference revived the auxiliary-variable inversion in a modern form (Hannig, 2009). IMs synthesize these lines by treating the auxiliary variable as the source of uncertainty and by predicting its unobservable realization with calibrated predictive random sets (PRSs) (Martin & Liu, 2013, 2015b).

The present review focuses on IMs and is written from the viewpoint that scientific uncertainty is more than sampling variability or aleatory uncertainty. It is also epistemic: uncertainty about fixed but unknown features of the world, about model adequacy, and about which inferential statements are warranted by the data. Under genuine prior information, epistemic uncertainty can be represented probabilistically. Under weak or absent prior information, a precise posterior distribution can create a false sense of confidence (Balch et al., 2019; Fraser, 2011; Fraser et al., 2016). IMs respond by using belief, plausibility, and, in recent work, possibility measures (Martin, 2019, 2026b), with validity providing the calibration guarantee needed for reliable uncertainty assessment. For readers unfamiliar with imprecise probabilities, the key point for scientific uncertainty quantification is that these measures retain a probability-like interpretation while allowing uncertainty to remain imprecise when the available information does not warrant a precise probability distribution.

Because IMs are less familiar than Bayesian posteriors, confidence intervals, or likelihood ratios, this review uses four running examples as a bridge between the abstract construction and practical inference. These examples reappear selectively in later sections so that the reader can see how the IM answer changes when one moves from a basic IM to conditional or marginal IM analyses in a validity-first and efficiency-driven fashion.

(1) **Normal models:** $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$. The known-$\sigma$ case gives the cleanest auxiliary pivot and shows how a symmetric PRS reproduces the usual normal confidence interval as an IM plausibility region. It will also be used to separate the calibrated success of the flat Jeffreys' prior for μ from the stronger claim that a formal prior represents objective ignorance. The unknown-σ case serves as the simplest example in which conditional IM reasoning and marginal IM reasoning both recover the familiar calibrated

inference for μ without requiring a joint prior for (μ, σ). It also shows why Jeffreys' priors can be delicate even in this familiar two-parameter model.

(2) **Binomial model:** $X \sim \text{Bin}(n, p)$. Following Clopper and Pearson (1934), we consider this example fundamental for developing statistical theory and methods because the inverse association is naturally set-valued, that is, a single auxiliary value can correspond to an interval of p values. That feature makes binomial inference a useful entry point to Dempster-Shafer's set-valued mappings and belief functions, and it shows why random sets and plausibility are more honest than a forced additive posterior-like probability in discrete problems (Dempster, 1967; Shafer, 1976).

(3) **Eight schools:** the data are the estimated effects of eight SAT coaching programs $y = (28, 8, -3, 7, -1, 1, 18, 12)$ with known standard errors $\sigma = (15, 10, 16, 11, 9, 11, 10, 18)$. In this classic example of Bayesian data analysis (Gelman et al., 2013), the usual hierarchical model is

$$y_j \mid \theta_j \sim N(\theta_j, \sigma_j^2), \quad \theta_j \mid \mu, \tau \sim N(\mu, \tau^2),\ j = 1, \ldots, 8. \tag{1}$$

For this eight-schools model, it is better for present purposes first to integrate out the latent school effects in (1) and obtain the marginal sampling model

$$y_j \overset{\text{ind}}{\sim} N(\mu, \tau^2 + \sigma_j^2), \quad j = 1, \ldots, 8. \tag{2}$$

Prior specification for the between-school standard deviation τ strongly affects the scientific story, especially the amount of pooling and the posterior mass near homogeneity $\tau = 0$ (Gelman, 2006; Gelman et al., 2013; Rubin, 1981). In this review the example appears as a constrained problem for $\tau \geq 0$ with μ fixed, illustrating elastic PRSs. Treating μ as a nuisance parameter would require a fuller conditional or marginal IM analysis, which is beyond the scope of this review and will be reported elsewhere.

(4) **Bivariate normal correlation:** (Y_{1i}, Y_{2i}), $i = 1, \ldots, n$, are independent bivariate normal observations with known zero means, known unit variances, and unknown correlation coefficient ρ. In this example, the sufficient statistic is two-dimensional while the target is one-dimensional. Simpler than the eight-schools example, this is a benchmark for conditional inference because no single global ancillary reduction is available. It motivates local conditional IMs (Martin & Liu, 2015a).

In addition to these running examples, the Behrens–Fisher problem is introduced later in the discussion of marginal IMs.

The article proceeds historically and conceptually as a review through examples. It starts in Section 2 with the background material for IMs: Section 2.1 reviews Bayesian prior specification and questions the probabilistic uncertainty interpretation of information-based priors, while Section 2.2 revisits fiducial inference and confidence intervals as prior-free attempts to recover post-data uncertainty. Section 3 develops the basic IM construction, reviews elastic PRSs for constrained problems, and then works through the running examples. Sections 4 and 5 then review two advanced IM developments for efficient inference: conditional IMs reduce auxiliary dimensions by conditioning, while marginal IMs target interest parameters in the presence of nuisance structure. Section 6 closes with a brief conclusion, points readers to several further unsettled topics, and sheds light on Jaynes' view of probability as the logic of science.

2. Bayes, fiducial inference, and confidence intervals

2.1. Bayes and the problem of prior specification

Bayes requires total probabilistic specification. The difficulty begins when formal priors are presented as objective representations of scientific ignorance. As the most important example, Jeffreys' rule chooses a density proportional to the square root of the Fisher information determinant (Jeffreys, 1946, 1948). Logically, Jeffreys' rule solves a particular formal problem: if ignorance could be represented by a density, and if that representation is required to be invariant under smooth changes of coordinates, then the Jeffreys density would be the natural answer. But formal ignorance does not by itself determine an additive probability distribution with calibrated scientific meaning (Martin & Liu, 2015b). Jaynes' maximum entropy principle (Jaynes, 1957, 2003) should be read similarly. Reference priors (Berger et al., 2009; Bernardo, 1979) are another principled formal construction, often target- and ordering-dependent, so their usefulness should likewise be separated from the stronger claim that they supply objective prior knowledge.

Although they often produce excellent answers, these priors always require care when interpreted as objective scientific uncertainty. The normal example shows both the success and the danger. If $X \sim N(\mu, \sigma^2)$ with known σ, the flat prior $\pi(\mu) \propto 1$ gives

$$\mu \mid x \sim N(x, \sigma^2)$$

for one observation, or $N(\bar{x}, \sigma^2/n)$ for a sample with the sample mean $\bar{x}$. The resulting central credible intervals coincide with the usual confidence intervals and with the fiducial intervals obtained from the pivot $(\bar{X} - \mu)/(\sigma/\sqrt{n})$. While this is a good interval procedure, it does not follow that a flat density is an objective probability distribution over μ, nor that all posterior probabilities computed from it have objective scientific meaning.

With unknown σ, one must distinguish two different formal priors. The full Jeffreys' prior for this two-parameter model, with parameters (μ, σ), is

$$\pi_J(\mu, \sigma) \propto \{\det I(\mu, \sigma)\}^{1/2} \propto \frac{1}{\sigma^2}.$$

The more popular prior is

$$\pi(\mu, \sigma) \propto \frac{1}{\sigma}$$

which is the right-Haar, independence Jeffreys, and reference prior for the usual location-scale analysis. It is this right-Haar prior that matches the standard t-distribution posterior for μ after integrating out σ. Again, the resulting one- and two-sided intervals have a compelling confidence or plausibility interpretation. But that success belongs to the pivot and its calibration, not to the claim that the full Jeffreys' prior σ^{-2}, or any other formal prior, is an objective probability representation of ignorance over (μ, σ).

For $p \in (0, 1)$ in the binomial example $X \sim \text{Bin}(n, p)$, Jeffreys' prior is $\text{Beta}(1/2, 1/2)$, while a uniform formal prior (Bayes, 1763) gives $\text{Beta}(1, 1)$. These choices lead to different posterior intervals, especially near $x = 0$ or $x = n$.

The bivariate normal correlation coefficient gives a different warning about whether formal priors can be consistent with Bayes as a rule for sequentially combining information. The parameter $\rho \in (-1, 1)$ is scalar, but the natural sufficient statistic has two components.

To see the ambiguity, transform the observations to

$$Y'_{1i} = 2^{-1/2}(Y_{1i} + Y_{2i}), \quad Y'_{2i} = 2^{-1/2}(Y_{1i} - Y_{2i}),$$

so that the two transformed samples are independent with

$$Y'_{1i} \sim N(0, 1 + \rho), \quad Y'_{2i} \sim N(0, 1 - \rho).$$

If the Y'_{1i}'s are regarded as the first data set, Jeffreys' rule for this first marginal model gives

$$\pi_+(\rho) \propto \frac{1}{1+\rho}.$$

A sequential Bayesian analysis then updates this prior by the likelihood for the Y'_{1i}'s and uses the resulting posterior as the prior when the Y'_{2i}'s become available; the final posterior is therefore proportional to the joint likelihood times $\pi_+(\rho)$. Reversing the order gives instead the first-stage Jeffreys' prior

$$\pi_-(\rho) \propto \frac{1}{1-\rho},$$

and the final posterior is proportional to the same joint likelihood times $\pi_-(\rho)$. A one-shot Jeffreys' calculation from the joint transformed model gives a third prior,

$$\pi_J(\rho) \propto \frac{(1+\rho^2)^{1/2}}{1-\rho^2}.$$

Thus, the sequential-updating route depends on which transformed component is processed first, while the joint-model route gives a different formal prior again.

The eight-schools example is even more revealing. The parameter τ controls heterogeneity. A prior on τ is not a harmless technicality. It determines how strongly the analysis shrinks school effects toward a common mean, and how much posterior mass is placed near homogeneity. The formal-prior ambiguity is similar to that in both the normal example and the bivariate normal correlation example: it arises when applying Jeffreys' or reference-prior reasoning to the marginal model for (μ, τ). Flat priors, half-Cauchy priors, and Jeffreys-type priors can therefore give noticeably different scientific narratives. This is not a failure of Bayes when the prior is genuine; it is a warning against calling formal priors objective.

The position taken here is therefore sharp but not anti-Bayesian. If one wants to do Bayes, one should be genuinely Bayesian and specify the prior as real scientific, subjective, or empirical. If one does not have such information, then IMs offer a way to quantify uncertainty without pretending that a formal prior has solved the problem. Information-based Bayes remains indispensable, especially for understanding probability as a disciplined logic of uncertain reasoning. But the scientific interpretation of its posterior probabilities must be earned, often through calibration, matching, or an IM plausibility interpretation, rather than assumed from the formal prior alone. This is close in spirit to Rubin's applied Bayesian position that frequency calculations can be essential for understanding, communicating, validating, and monitoring Bayesian analyses (Rubin, 1984).

2.2. Fiducial inference and confidence intervals

Historically, Fisher's fiducial argument preceded Neyman's mature confidence theory. Expositionally, however, it is helpful to begin with familiar confidence intervals, because Neyman's construction gives the mathematically clean interpretation of many fiducial intervals. Neyman's 1941 paper is explicitly titled 'Fiducial argument and the theory of confidence intervals', and the later confidence-distribution literature also treats many fiducial distributions as confidence distributions in regular one-parameter settings (Neyman, 1941; Schweder & Hjort, 2016; Xie & Singh, 2013).

In Neyman's theory, probability attaches to the random procedure before the data are observed (Neyman, 1937). A $100(1-\alpha)\%$ confidence procedure $C(X)$ satisfies

$$\mathsf{P}_{X|\theta}\{\theta \in C(X)\} \geq 1-\alpha$$

for each θ in the parameter space. Once $X = x$ is observed, $C(x)$ is a fixed set. The classical theory does not say that the fixed θ lies in this fixed set with probability $1-\alpha$. The gain is conceptual clarity; the cost is a limited post-data evidential interpretation.

Fisher wanted more for situation-specific probabilistic inference. In the known-variance normal model,

$$\bar{X} = \mu + \frac{\sigma}{\sqrt{n}}Z, \quad Z \sim N(0,1),$$

the observed value $\bar{x}$ of the sample mean $\bar{X}$ gives the formal inversion

$$\mu = \bar{x} - \frac{\sigma}{\sqrt{n}}Z.$$

If one continues to regard Z as standard normal after $\bar{X} = \bar{x}$ is observed, then μ receives the fiducial distribution $N(\bar{x}, \sigma^2/n)$. Its central intervals match the usual confidence intervals. In this example, Fisher's fiducial answer, Neyman's confidence interval, and the flat-prior Bayesian interval agree numerically. The unknown-variance normal model gives the next classical case. The pivot

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}, \tag{3}$$

where S^2 is the unbiased sample variance as the estimator of σ^2, leads to the usual t interval for μ. It can be read as a confidence interval, a fiducial interval, or a formal Bayesian credible interval under the right-Haar/reference prior $\pi(\mu,\sigma) \propto 1/\sigma$.

The binomial case shows why the above equivalence breaks down. If $X \sim \mathrm{Bin}(n,p)$, a natural auxiliary representation uses $U \sim \mathrm{Unif}(0,1)$ and

$$X = F_p^{-1}(U),$$

where F_p is the binomial distribution function. After observing $X = x$, the inversion is not a single value of p for each u; it is generally a set determined by

$$F_p(x-1) \leq u < F_p(x). \tag{4}$$

Clopper and Pearson (1934) explicitly presented their construction in confidence or fiducial language for this binomial problem. Its intervals are exact but conservative, and discreteness exposes the gap between calibrated set inference and precise posterior-like probability.

The bivariate normal correlation coefficient and eight schools expose a second gap, this time caused by conditioning. In the bivariate normal correlation, for example, the transformed sums

$$S_1 = \frac{1}{2}\sum_{i=1}^{n}(Y_{1i} + Y_{2i})^2, \quad S_2 = \frac{1}{2}\sum_{i=1}^{n}(Y_{1i} - Y_{2i})^2$$

serve as sufficient statistics and satisfy $S_1 = (1+\rho)U_1$ and $S_2 = (1-\rho)U_2$, where U_1, U_2 are independent χ^2_n variables. It is tempting to make fiducial inference by conditioning on an identity such as $S_1/U_1 + S_2/U_2 = 2$. Although this identity eliminates ρ, it does not identify a nontrivial observed auxiliary feature through a separable relation $H(S_1, S_2) = \eta(U_1, U_2)$, with both maps independent of ρ. Geometrically, for fixed observed $s = (s_1, s_2)$, the identity restricts the auxiliary value to the data-dependent compatibility curve

$$\mathcal{C}_s = \left\{(u_1, u_2) \in (0,\infty)^2 : \frac{s_1}{u_1} + \frac{s_2}{u_2} = 2\right\}.$$

These curves can intersect as s varies, so they do not form the level sets of a common auxiliary statistic whose value has been observed. Indeed, for this association, requiring a smooth η to remain constant along the auxiliary motion induced by every possible ρ forces η itself to be constant. Thus, the identity supplies a compatibility constraint, but does not by itself justify conditioning the auxiliary law on that constraint or guarantee the validity of the resulting inference. Classical likelihood theory handles this through ancillary or approximate conditional arguments, including the $r^\star$ machinery of Barndorff-Nielsen (1986), Fraser (1990), and Reid (1995). The IM response is to ask a more primitive question: *which lower-dimensional auxiliary quantity can be predicted validly after the observed auxiliary features have been conditioned on?* Conditional IMs answer this question later in Section 4.

Nevertheless, fiducial reasoning has remained a source of inspiration, and the historical study of Zabell (1992) serves as an essential guide. Fisher identified a real scientific need for posterior-like reasoning without arbitrary priors. But the fiducial argument did not become a general coherent probability calculus. Fraser's structural equations (Fraser, 1968), Dempster-Shafer theory (Dempster, 1967; Shafer, 1976), and generalized fiducial inference (Hannig, 2009; Hannig et al., 2016) are three later developments inspired by this unfinished ambition. These are crucial predecessors of IMs. The IM contribution is to keep the auxiliary-variable source of uncertainty but change the order of operations and seek ways to improve efficiency through conditioning and marginalization in the auxiliary space.

3. Inferential models

3.1. The basic IM construction

The IM framework begins with an association

$$X = a(\theta, U), \quad U \sim \mathsf{P}_U, \tag{5}$$

where X is observable, θ is unknown, and U is an auxiliary variable with known distribution. After $X = x$ is observed, there exists an unobserved auxiliary value $u^\star$ consistent with the true parameter and the observed data. The inferential problem is to reason about θ through uncertainty about $u^\star$.

The distinction from fiducial reasoning is the central point. Fiducial methods transfer the auxiliary variable to the parameter space first, producing a data-dependent parameter distribution, and then apply probability calculus in the parameter space. IMs carry out the probability operation before that transfer (Liu & Martin, 2015); see also Barnard (1995). They predict $u^\star$ in the well-defined auxiliary-variable probability space by a valid PRS, and only after this prediction do they combine the result with the observed association.

Procedurally, the basic IM has three steps (Martin & Liu, 2013, 2015b).

(1) **Association.** Specify the association (5) that preserves the sampling model of X. For fixed x and u, this defines the compatible parameter set

$$\Theta_x(u) = \{\theta : x = a(\theta, u)\}.$$

(2) **Prediction.** Predict the unobserved $u^\star$ by a PRS $\mathcal{S} \sim \mathsf{P}_\mathcal{S}$ on the auxiliary space $\mathbb{U}$. The mathematical calibration requirement for $\mathcal{S}$ is imposed before any transfer to the parameter space.

(3) **Combination.** Map the predictive random set to the parameter space:

$$\Theta_x(\mathcal{S}) = \bigcup_{u \in \mathcal{S}} \Theta_x(u).$$

For an assertion $A \subseteq \Theta$, define the belief and plausibility

$$\text{bel}_x(A) = \mathsf{P}_\mathcal{S}\{\Theta_x(\mathcal{S}) \subseteq A\}, \quad \text{pl}_x(A) = \mathsf{P}_\mathcal{S}\{\Theta_x(\mathcal{S}) \cap A \neq \emptyset\}.$$

Thus, belief measures definite support for A; plausibility measures the extent to which A has not been ruled out. Unlike a precise additive probability, belief and plausibility generally provide lower and upper assessments of support, with $\text{bel}_x(A) \leq \text{pl}_x(A)$. The gap $\text{pl}_x(A) - \text{bel}_x(A)$ reflects inferential imprecision rather than additional randomness. For point assertions, these assessments can often be summarized by the plausibility contour $\theta \mapsto \text{pl}_x(\{\theta\})$, whose level sets yield plausibility regions and which provides the graphical summaries used below in Section 3.3.

This basic IM is oriented toward validity first. Formally, the calibration logic can be stated as two definitions and a theorem.

Definition 3.1 (Valid PRS): For a PRS $\mathcal{S} \sim \mathsf{P}_\mathcal{S}$ on the auxiliary space $\mathbb{U}$, define the PRS contour

$$\gamma_\mathcal{S}(u) = \mathsf{P}_\mathcal{S}\{\mathcal{S} \ni u\}, \quad u \in \mathbb{U}. \tag{6}$$

The PRS is *valid* for predicting the auxiliary value if

$$\mathsf{P}_U\{\gamma_\mathcal{S}(U) \leq \alpha\} \leq \alpha, \quad 0 < \alpha < 1, \tag{7}$$

where $U \sim \mathsf{P}_U$. Equivalently, $\gamma_\mathcal{S}(U)$ is stochastically no smaller than a $\text{Unif}(0,1)$ random variable. In miss-probability form, with

$$Q_\mathcal{S}(u) = \mathsf{P}_\mathcal{S}\{\mathcal{S} \not\ni u\} = 1 - \gamma_\mathcal{S}(u),$$

validity is

$$\mathsf{P}_U\{Q_\mathcal{S}(U) \geq 1 - \alpha\} \leq \alpha, \quad 0 < \alpha < 1.$$

Thus, a large miss probability can occur only on a small P_U-set of auxiliary values.

Definition 3.2 (Valid posterior assertion plausibility): For an assertion $A \subseteq \Theta$, the IM procedure gives the posterior assertion plausibility

$$\text{pl}_x(A) = \mathsf{P}_{\mathcal{S}}\{\Theta_x(\mathcal{S}) \cap A \neq \emptyset\}.$$

This plausibility is *valid* for assertion A if

$$\sup_{\theta \in A} \mathsf{P}_{X|\theta}\{\text{pl}_X(A) \leq \alpha\} \leq \alpha, \quad 0 < \alpha < 1. \tag{8}$$

Equivalently, when A is true, $\text{pl}_X(A)$ is stochastically no smaller than a $\text{Unif}(0,1)$ random variable. The IM is valid if (8) holds for all assertions $A \subseteq \Theta$.

The basic IM validity theorem of Martin and Liu (2013) links auxiliary prediction to assertion-wise calibration.

Theorem 3.3 (Validity theorem): *If the PRS $\mathcal{S}$ is valid in the sense of Definition* 3.1 *and $\Theta_x(\mathcal{S})$ is non-empty with $\mathsf{P}_{\mathcal{S}}$-probability one for each x, then the IM obtained by the association, prediction, and combination steps is valid in the sense of Definition* 3.2 *for every assertion A. Consequently, the plausibility region*

$$\Pi_x(\alpha) = \{\theta : \text{pl}_x(\{\theta\}) > \alpha\}$$

satisfies

$$\inf_{\theta \in \Theta} \mathsf{P}_{X|\theta}\{\theta \in \Pi_X(\alpha)\} \geq 1 - \alpha,$$

so the fixed-data plausibility scale has a direct frequentist calibration.

Remark 3.1 (Nested PRSs and possibility measures): An important result in the original IM developments is that, with regard to efficiency, there is no need to consider random sets $\mathcal{S}$ that are not nested. When $\mathcal{S}$ is nested, typically $\Theta_x(\mathcal{S})$ is nested too. In such cases, the belief function is consonant and the corresponding plausibility function has the properties of a possibility measure. This is important for several reasons. First, it is only for possibility measures that the contour function can be treated as a primitive, i.e., the object that fully determines the IM output. Having such a simple point function, rather than a set function, that determines the IM output is valuable, like how countable additivity allows for probabilities to be determined by a density/mass function. For general belief functions, the 'contour' is only a derivative object, not a primitive, so the full/complex set function is required to describe the IM solution. Second, it explains the recent alternative formulation that directly defines the IM as a possibility measure (see Martin, 2026b, and references therein). That is, if the resulting IM will be a possibility measure anyway, it can be more expedient to define the IM directly as such. The auxiliary-variable-based predictive random set then becomes implicit rather than explicit, with its essential role preserved through the calibration condition that guarantees validity. For non-consonant belief functions, by contrast, singleton plausibilities do not determine the full set function, so the contour alone is insufficient to represent the IM output.

3.2. Elastic PRSs for constrained problems

The basic construction above assumes that $\Theta_x(\mathcal{S})$ is non-empty with probability one. This assumption can fail, for example, in constrained problems. Suppose the parameter is known to lie in a proper subset $C \subseteq \Theta$, such as a nonnegative mean, a variance component, or a Poisson signal rate with known background. For fixed data x, the constraint maps back to the auxiliary space through

$$U_{C,x} = \bigcup_{\theta \in C} \{u : x = a(\theta, u)\}.$$

If the realized PRS $\mathcal{S}$ misses $U_{C,x}$, then the corresponding parameter random set is incompatible with the constraint. These are referred to as conflict cases.

One solution to the conflict-case problem is to condition away the conflict cases, as in Dempster's rule of combination. This is valid in many settings, but it can be inefficient because it discards informative conflict cases or redistributes more predictive mass than is needed. Elastic PRSs were introduced by Ermini Leaf and Liu (2012) to remove conflict cases more locally. Start with a valid baseline PRS $\mathcal{S}_0$ and embed it in an increasing family

$$\mathcal{S}_e, \quad 0 \leq e \leq 1,$$

where $\mathcal{S}_0$ is the original PRS and $\mathcal{S}_1$ is large enough to guarantee compatibility. For each realized auxiliary prediction, choose the smallest stretching amount

$$\widehat{e} = \inf\{e : \mathcal{S}_e \cap U_{C,x} \neq \emptyset\},$$

and combine with

$$\Theta_x(\mathcal{S}_{\widehat{e}}) \cap C.$$

Thus, the PRS stretches only until it intersects the auxiliary values compatible with the constraint. The elastic step is data-dependent, but it is not an arbitrary posterior truncation. It is carried out in the auxiliary space, where validity is controlled, and then transferred to the constrained parameter space.

The key property is that this minimal stretching preserves calibration while improving efficiency. Ermini Leaf and Liu (2012) show that the resulting elastic belief function is valid for assertions $A \subseteq C$ and is typically more efficient than Dempster's rule of combination. Elastic PRSs are important conceptually because constraints are common in scientific inference. Variance components, signal rates, scale parameters, probabilities, and physical effects often live on restricted spaces. Bayesian methods usually encode such restrictions by truncating or reshaping a prior.

3.3. Illustrative examples

The following selected numerical illustrations pair fixed-data plausibility curves with simulation diagnostics for validity. In the diagnostics, the plausibility assigned to the true parameter value is computed repeatedly under the model and compared with the Unif(0, 1) benchmark; exact continuous IMs should track the diagonal, while discrete examples often unavoidably appear conservative. The diagnostic plots are shown in Figures 1, 2, and 3 for the normal, binomial, and eight-schools examples that are discussed below.

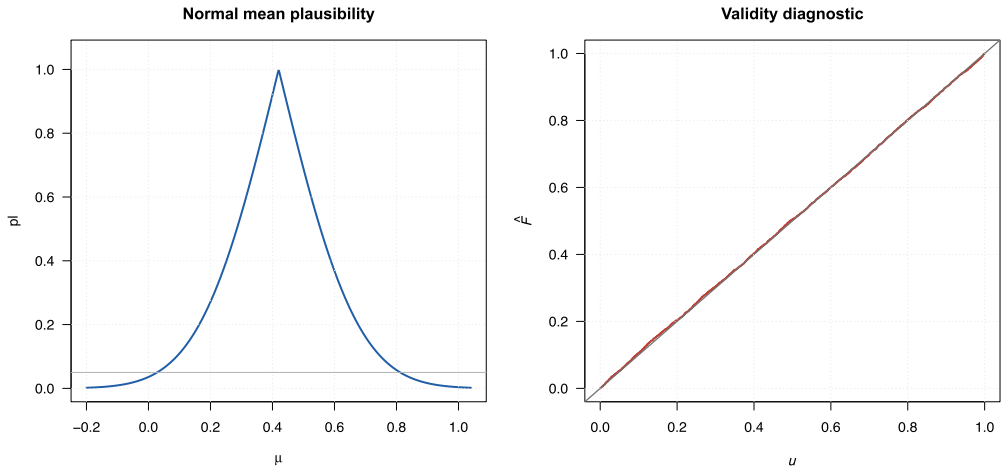


Figure 1. Normal mean example. The left panel shows the plausibility curve for μ when $n = 25$, $\sigma = 1$, and $\bar{x} = 0.42$. The right panel is a validity diagnostic based on 5000 simulations under $\mu = 0$; the empirical distribution of $\text{pl}_x(0)$ closely follows the uniform diagonal.

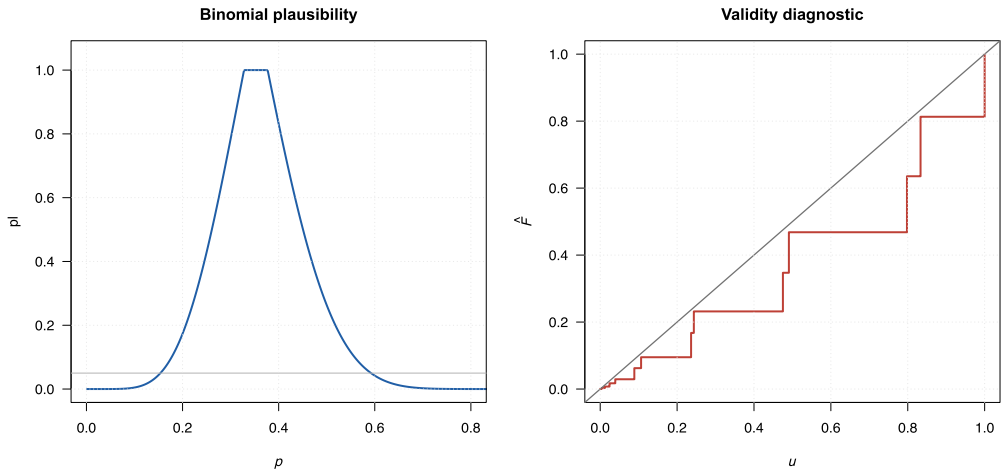


Figure 2. Binomial example. The left panel shows an exact two-sided plausibility curve for p when $n = 20$ and $x = 7$. The right panel shows a validity diagnostic based on 5000 simulations under $p = 0.35$; discreteness makes the distribution conservative, so the empirical curve lies mostly below the uniform diagonal.

In the known-variance normal example, the conditional-IM reduction (see Section 4) gives the association for the reduced sufficient statistic

$$\bar{X} = \mu + \frac{\sigma}{\sqrt{n}} Z, \quad Z \sim N(0,1).$$

An estimation-oriented default symmetric predictive random set is

$$\mathcal{S} = \{z : |z| \leq |Z'|\}, \quad Z' \sim N(0,1).$$

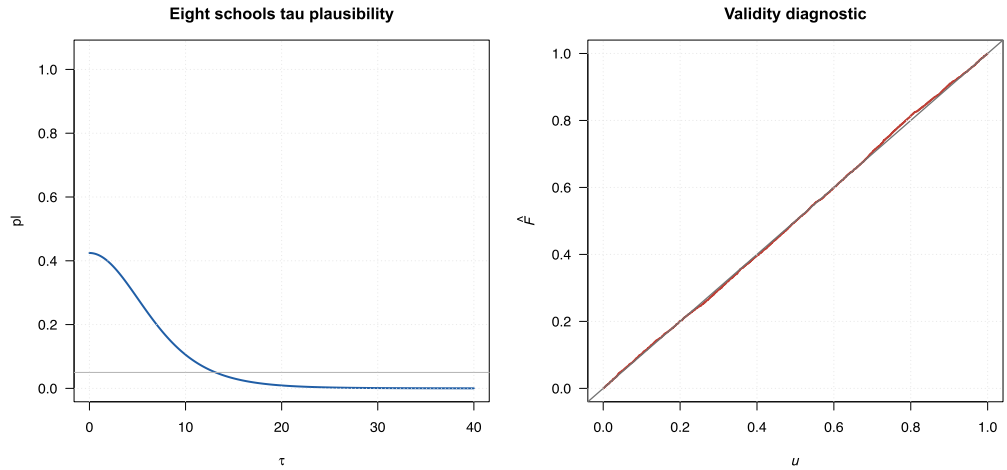


Figure 3. Eight-schools heterogeneity example with $\mu = 8$ fixed. The left panel shows the plausibility contour for the constrained parameter $\tau \geq 0$ using the marginal model after integrating out the school effects. Under the elastic-PRS construction, the contour takes the value 1 at the boundary $\tau = 0$ and has an immediate downward jump to approximately 0.42 for $\tau > 0$; this discontinuity reflects the stretching of conflicting predictive random sets just enough to restore compatibility with the constraint. The estimated probability that the baseline central PRS is in conflict and therefore requires elastic stretching is 0.576. The right panel is a validity diagnostic based on 5000 simulations under $\tau = 10$.

For the singleton assertion $\{\mu\}$, the plausibility contour is

$$\mathrm{pl}_x(\mu) = 1 - \left|2\Phi\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right) - 1\right|.$$

The plausibility set $\{\mu : \mathrm{pl}_x(\mu) > \alpha\}$ is exactly the usual two-sided $100(1-\alpha)\%$ normal interval. Thus, the familiar interval is not discarded. It is reinterpreted as a calibrated plausibility region arising from valid prediction of the auxiliary variable and giving the desired interpretation of situation-specific inference.

For the unknown-variance normal problem, after efficiency operations (see Sections 4 and 5), one can work with (3). Using the same symmetric predictive-random-set idea on the t auxiliary scale gives

$$\mathrm{pl}_x(\mu) = 1 - \left|2F_{t_{n-1}}\left(\frac{\bar{x} - \mu}{s/\sqrt{n}}\right) - 1\right|,$$

and the plausibility regions coincide with the usual t intervals. Again, this agreement does not say that all formal posterior probabilities are objective. It says that the interval has a calibrated auxiliary-variable predictive inference justification.

For the binomial model, the association $X = F_p^{-1}(U)$ with $U \sim \mathrm{Unif}(0,1)$ gives (4) and

$$\Theta_x(u) = \{p : F_p(x-1) \leq u < F_p(x)\}.$$

Because $\Theta_x(u)$ is set-valued and the distribution is discrete, a singleton fiducial transfer is unnatural. Like Dempster-Shafer, an IM keeps the set-valued nature visible. Predicting $u^\star$ by a valid PRS yields plausibility regions closely related to exact binomial tests and Clopper-Pearson intervals, but the output is belief and plausibility rather than a forced confidence interval for p.

For the eight-schools model illustration, take μ to be known and focus on $\tau \geq 0$. This turns the example into a clean constrained-parameter problem. An auxiliary association is

$$Y_j = \mu + (\tau^2 + \sigma_j^2)^{1/2} Z_j, \quad Z = (Z_1, \ldots, Z_8) \sim N_8(0, I).$$

For a candidate τ, the auxiliary value required by the observation is

$$z_\tau(y) = \left(\frac{y_1 - \mu}{(\tau^2 + \sigma_1^2)^{1/2}}, \ldots, \frac{y_8 - \mu}{(\tau^2 + \sigma_8^2)^{1/2}} \right).$$

Given a PRS $\mathcal{S}$ for Z, the value τ is plausible if $z_\tau(y) \in \mathcal{S}$. Without further reducing dimension via conditional IMs, for example, a central nested PRS based on $h(z) = \sum_j z_j^2$ gives a plausibility contour of the form

$$\text{pl}_y(\tau) = 1 - \left| 2F_{\chi_8^2}\{h(z_\tau(y))\} - 1 \right|, \quad \tau \geq 0,$$

where $F_{\chi_8^2}$ is the χ_8^2 distribution function. This two-sided form treats both unusually large and unusually small standardized residual sums as atypical. The constraint $\tau \geq 0$ is not cosmetic: if the combined random set $\{\tau \geq 0 : z_\tau(y) \in \mathcal{S}\}$ is empty for a realized PRS, the elastic PRS construction stretches the auxiliary prediction only enough to hit the curve $\{z_\tau(y) : \tau \geq 0\}$. For the observed data with $\mu = 8$, $h(z_0(y)) = 4.713$, $\text{pl}_y(0) = 0.424$, and the probability that the baseline central PRS misses the constrained auxiliary curve is $1 - \sup_{\tau \geq 0} \text{pl}_y(\tau) = 0.576$. This is the probability that the elastic adjustment is activated in this illustration.

This conflict probability is not itself a point on the plausibility curve. The curve reports, for each fixed τ, how plausible that value is after combining the data, the PRS, and the constraint. The conflict probability is a diagnostic of the baseline PRS before elastic repair: it measures how often the auxiliary prediction fails to hit any point on the constrained curve $\{z_\tau(y) : \tau \geq 0\}$. Thus, it is naturally reported as a separate additional annotation or diagnostic. The case with unknown μ introduces a nuisance parameter and is better treated as both a conditional IM and a marginal IM problem. A full treatment thus requires substantial space and will be reported elsewhere.

The bivariate normal correlation example is not developed here because its main role is to motivate local conditional IMs, where it will be treated explicitly. The philosophical lesson across the examples is the same. The sampling model supplies an auxiliary source of randomness. Unlike fiducial reasoning, IMs keep the randomness in its native auxiliary space, calibrate the prediction there, and then transfer the resulting set-valued uncertainty to the parameter space. This is why IMs can preserve Fisher's post-data ambition and Neyman's calibration requirement at the same time.

4. Conditional IMs

Conditional IMs (Martin & Liu, 2015a) are the first major efficiency refinement of the basic IM construction. A baseline association may contain many auxiliary coordinates even when the parameter or assertion of interest is low-dimensional. Predicting all of those coordinates is valid but wasteful, and wasteful prediction produces inefficient plausibility. The conditional IM idea is to identify auxiliary features that have effectively been observed, condition on them, and predict only the remaining unobserved auxiliary feature. Classical reviews

of ancillary conditioning make clear why this matters: conditioning can define the relevant subset of the sample space and can make inference more efficient (Fraser, 2004; Ghosh et al., 2010; Reid, 1995). The difference is that conditional IMs carry out conditioning in the auxiliary-variable space.

4.1. Regular conditional IMs

Start from a baseline association (5). A regular conditional IM is available when the association can be rewritten, after one-to-one transformations of data and auxiliary variables, as

$$h_1(X) = a_1\{\phi_1(U), \theta\}, \quad h_2(X) = a_2\{\phi_2(U)\}, \tag{9}$$

where the second equation is free of θ. The observed value $h = h_2(x)$ reveals the corresponding feature of the auxiliary variable. Therefore, instead of predicting the full U, the conditional IM predicts

$$V = \phi_1(U)$$

under the conditional distribution of V given $a_2\{\phi_2(U)\} = h$. The association used for inference is the lower-dimensional conditional association

$$h_1(x) = a_1(v, \theta), \quad V \sim \mathsf{P}_{\phi_1(U)|a_2\{\phi_2(U)\}=h}.$$

The A/P/C steps that produce inferential assessments are then the same as before: invert this conditional association, predict the unobserved $V^\star$ using a predictive random set on the conditional auxiliary space and combine the results.

The important change is that validity is now conditional. If the predictive random set $\mathcal{S}$ is conditionally valid under the conditional law at h, then the resulting plausibility is calibrated on the relevant subset $\{x : h_2(x) = h\}$. In particular, for a true assertion A,

$$\mathsf{P}_\theta\{\mathrm{pl}_x(A) \leq \alpha \mid h_2(X) = h\} \leq \alpha, \quad 0 < \alpha < 1,$$

for $\theta \in A$ under the usual non-empty-focal-set condition. This is stronger and more interpretable than an unconditional guarantee when the conditioning statistic really is part of the observed auxiliary information.

Regular conditional IMs clarify the relation between IMs, sufficiency, ancillarity, and Bayes. If $h_1(X)$ is sufficient and the conditional law of $\phi_1(U)$ does not depend on h, then the conditional reduction reproduces the familiar instruction to work with the sufficient statistic. If $h_2(X)$ is ancillary, then validity is assessed conditionally on the ancillary value, matching Fisher's relevant-subset idea. From the IM viewpoint, both are special cases of a more primitive operation: combine information by conditioning on observed auxiliary features before prediction. This is also why conditional IMs can be viewed as a prior-free counterpart to Bayesian information aggregation, but without requiring a prior distribution on θ.

Here, regularity requires that the same conditioning map applies across the parameter space, and the conditioning Equation (9) must separate data from auxiliary variables in a parameter-free way. Many textbook models have this form after a sensible transformation. But there are important models in which the transformation (9) does not exist. In such cases, no single global conditioning feature is available from (5), and one can consider alternative specifications of (5) or local conditional IMs.

4.2. Local conditional IMs

Local conditional IMs (Martin & Liu, 2015a) were introduced for precisely those nonregular cases. The idea is to assess each singleton assertion $\{\theta_0\}$ using a conditional association localized at θ_0. For a fixed candidate value, one seeks transformations

$$T_{\theta_0}(X) = b_{\theta_0}\{\tau_{\theta_0}(U), \theta_0\}, \quad H_{\theta_0}(X) = \eta_{\theta_0}(U),$$

at least locally around θ_0, where $H_{\theta_0}(X)$ is the observed auxiliary feature to condition on and $\tau_{\theta_0}(U)$ is the reduced auxiliary variable to be predicted. The plausibility of θ_0 is computed from the conditional distribution

$$\tau_{\theta_0}(U) \mid \eta_{\theta_0}(U) = H_{\theta_0}(x).$$

Then the local plausibility region is assembled as

$$\{\theta : \mathrm{pl}_{x,\theta}(\{\theta\}) > \alpha\},$$

where the subscript reminds us that the conditional association used to test θ or compute $\mathrm{pl}_x(\{\theta\})$ is localized at that same value.

This may look unusual at first because the conditioning law changes with the candidate value. But that is exactly what prevents an invalid fiducial shortcut. Local conditional IMs pay a small price in notation and computation to preserve the essential IM ordering: first identify the auxiliary feature still needing prediction, then predict it validly, and only then transfer uncertainty to the parameter assertion.

The validity statement is local but still meaningful. If, at θ_0, the conditional PRS is valid under the conditional law used by the localized association, then

$$\mathsf{P}_{\theta_0}\{\mathrm{pl}_{X,\theta_0}(\{\theta_0\}) \leq \alpha \mid H_{\theta_0}(X) = h\} \leq \alpha.$$

Thus, a local conditional IM supplies a calibrated plausibility scale for each candidate value, even when no global ancillary or sufficient reduction is available.

4.3. A differential geometric formulation

The construction can be described geometrically (Liu, 2026). For a smooth association $x = a(\theta, u)$, fixing x and varying θ traces a curve or surface through the auxiliary space. Assume, locally, that the derivative (square matrix) $D_u a(\theta, u)$ is nonsingular. Differentiating the identity $x = a(\theta, u_{x,\theta})$ with x fixed gives the parameter-motion vector fields

$$G_j^{\theta}(u) \equiv \frac{\partial u_{x,\theta}}{\partial \theta_j} = -\{D_u a(\theta, u)\}^{-1} \partial_{\theta_j} a(\theta, u), \quad j = 1, \ldots, p.$$

At a candidate value θ_0, a local conditioning map η_{θ_0} should be insensitive to these motions. In differential form it solves

$$D\eta_{\theta_0}(u) G_j^{\theta_0}(u) = 0, \quad j = 1, \ldots, p. \tag{10}$$

Thus, local conditioning variables are first integrals of the so-called *distribution* in differential geometry

$$\mathcal{D}_{\theta_0}(u) = \mathrm{span}\{G_1^{\theta_0}(u), \ldots, G_p^{\theta_0}(u)\}$$

on the auxiliary space.

This is the differential formulation behind local conditional IMs. It includes the class of regular conditional IMs as a special case. A precise mathematical theory is developed in Liu (2026).

4.4. Illustrative examples

4.4.1. Normal location and scale

The normal location model shows regular conditioning in its simplest form. With $X_i = \theta + U_i$ and $U_i \overset{\text{i.i.d.}}{\sim} N(0,1)$,

$$\bar{X} = \theta + \bar{U}, \quad X_i - \bar{X} = U_i - \bar{U},\ i = 1, \ldots, n.$$

The residual vector is an observed auxiliary feature, and $\bar{U}$ is independent of the residuals. The conditional IM therefore predicts only $\bar{U} \sim N(0, 1/n)$, giving the same plausibility contour as the usual normal pivot. This is a regular conditional IM because the conditioning feature does not depend on θ.

The unknown-variance normal model illustrates how conditional and marginal thinking meet. The details are discussed in Section 5.

4.4.2. Binomial model

This example connects directly to Dempster's 1960s innovation of lower and upper probabilities. Instead of starting from the scalar inverse-cdf representation, take

$$U_1, \ldots, U_n \overset{\text{i.i.d.}}{\sim} \text{Unif}(0,1), \quad X_i = 1_{\{U_i \le p\}}, \quad X = \sum_{i=1}^{n} X_i.$$

After observing only $X = x$, the individual labels of the uniforms carry no information about p. If $U_{(1)} \le \cdots \le U_{(n)}$ are the ordered uniforms, with the conventions $U_{(0)} = 0$ and $U_{(n+1)} = 1$, then the auxiliary values compatible with the count determine the random interval

$$\Gamma_x(U) = \{p : U_{(x)} \le p \le U_{(x+1)}\},$$

up to endpoint conventions that are immaterial under the continuous auxiliary law. Dempster's multivalued mapping idea assigns lower and upper probabilities by

$$\text{bel}_x(A) = \mathsf{P}\{\Gamma_x(U) \subseteq A\}, \quad \text{pl}_x(A) = \mathsf{P}\{\Gamma_x(U) \cap A \neq \emptyset\},$$

for assertions $A \subseteq [0,1]$ (Dempster, 1966, 1967, 1968). For example,

$$\begin{aligned} \text{pl}_x\{p \le p_0\} &= \mathsf{P}\{U_{(x)} \le p_0\} = \mathsf{P}_{p_0}(X \ge x), \\ \text{pl}_x\{p \ge p_0\} &= \mathsf{P}\{U_{(x+1)} \ge p_0\} = \mathsf{P}_{p_0}(X \le x). \end{aligned}$$

Thus, the ordered-uniform construction recovers the exact binomial tail probabilities as plausibilities, while retaining the set-valued inverse image forced by discreteness.

In contrast, the IM solution can be made even more direct by using the probability integral transform for the observed count itself. Let $U \sim \text{Unif}(0,1)$ and generate

$$X = F_p^{-1}(U),$$

where F_p is the binomial cdf. After observing $X = x$, a candidate p is compatible with an auxiliary value u exactly when

$$F_p(x-1) < u \leq F_p(x),$$

with the usual convention $F_p(-1) = 0$. Thus, inference for p can be based on a one-dimensional auxiliary variable rather than on Dempster's ordered-uniform endpoints. If $\mathcal{S}$ is a valid PRS for U, then the induced random set

$$\Theta_x(\mathcal{S}) = \{p : \mathcal{S} \cap (F_p(x-1), F_p(x)] \neq \emptyset\}$$

gives calibrated belief and plausibility for assertions about p. This preserves Dempster's set-valued insight for discreteness, but it locates the IM prediction problem on the scalar auxiliary scale tied directly to the observed count.

4.4.3. Bivariate normal correlation

The bivariate normal correlation coefficient is the cleanest illustration of the local procedure because the usual sufficiency reduction leaves one extra auxiliary dimension. With the notation of Section 2.2, the reduced association is

$$S_1 = (1+\rho)U_1, \quad S_2 = (1-\rho)U_2, \quad U_1, U_2 \overset{\text{ind}}{\sim} \chi^2_n,$$

where $-1 < \rho < 1$. Write $s = (s_1, s_2)$ for the observed sufficient statistic. For a fixed candidate value ρ, the auxiliary value required by the observation is

$$u_{s,\rho} = \left(\frac{s_1}{1+\rho}, \frac{s_2}{1-\rho}\right).$$

The parameter-motion vector field induced by varying ρ while keeping s fixed is therefore

$$G^\rho(u) = \frac{\partial u_{s,\rho}}{\partial \rho} = \left(-\frac{u_1}{1+\rho}, \frac{u_2}{1-\rho}\right)^\top.$$

Following (10), a local conditioning coordinate η_ρ must satisfy

$$-\frac{u_1}{1+\rho}\frac{\partial \eta_\rho}{\partial u_1}(u) + \frac{u_2}{1-\rho}\frac{\partial \eta_\rho}{\partial u_2}(u) = 0. \tag{11}$$

A corresponding first integral is

$$\eta_\rho(u) = (1+\rho)\log u_1 + (1-\rho)\log u_2,$$

since substituting this derivative in (11) gives $-1+1=0$. Hence, the observed local conditioning value is

$$\begin{aligned} h_\rho(s) &= \eta_\rho(u_{s,\rho}) \\ &= \log(s_1 s_2) + \rho\log(s_1/s_2) - \{(1+\rho)\log(1+\rho) + (1-\rho)\log(1-\rho)\}. \end{aligned}$$

This is the promised local version of $H_{\theta_0}(X) = \eta_{\theta_0}(U)$: the conditioning coordinate is fixed only after the candidate value has been fixed.

The complementary coordinate can be chosen as

$$\tau(U) = \log(U_1/U_2).$$

Then

$$T(S) = \log(S_1/S_2) = z(\rho) + V, \quad z(\rho) = \log\{(1+\rho)/(1-\rho)\}, \quad V = \tau(U).$$

Thus, to assess the singleton assertion $\{\rho\}$, the local conditional IM predicts V under the conditional law

$$V = \log(U_1/U_2) \mid \eta_\rho(U) = h_\rho(s).$$

This is exactly the A-step of the local construction: condition on the observed auxiliary feature and retain only the scalar auxiliary variable still needed to check compatibility with the candidate ρ.

The conditional density follows by an explicit change of variables. Let

$$v = \log(u_1/u_2), \quad h = (1+\rho)\log u_1 + (1-\rho)\log u_2.$$

Solving gives

$$\log u_1 = \frac{h + (1-\rho)v}{2}, \quad \log u_2 = \frac{h - (1+\rho)v}{2}.$$

Since U_1 and U_2 are independent χ^2_n variables, their joint density is proportional to

$$(u_1 u_2)^{n/2-1} \exp\{-(u_1+u_2)/2\}.$$

The Jacobian for $(v,h) \mapsto (u_1,u_2)$ is $(u_1u_2)/2$, and

$$u_1 u_2 = \exp(h - \rho v), \quad \frac{u_1+u_2}{2} = \cosh(v/2)\exp\{(h-\rho v)/2\}.$$

Therefore, conditional on $h = h_\rho(s)$, the density of V is

$$f_{\rho,h_\rho}(v) \propto \exp\left\{-\frac{n\rho v}{2} - \cosh(v/2)\exp\left(\frac{h_\rho(s) - \rho v}{2}\right)\right\}, \quad v \in \mathbb{R}. \tag{12}$$

Let F_{ρ,h_ρ} denote the distribution function corresponding to (12). For the observed s, the auxiliary value required by the candidate ρ in the conditional association is

$$c_\rho(s) = T(s) - z(\rho) = \log(s_1/s_2) - \log\{(1+\rho)/(1-\rho)\}.$$

Equivalently, $F_{\rho,h_\rho}\{c_\rho(s)\}$ is the scalar uniform auxiliary value associated with the candidate ρ after conditioning. Using the usual symmetric nested PRS on this scalar uniform scale gives the singleton plausibility contour

$$\mathrm{pl}_s(\rho) = 1 - \left|2F_{\rho,h_\rho}\{c_\rho(s)\} - 1\right|, \quad -1 < \rho < 1.$$

Geometrically, this derivation also shows why the example is not regular. A regular conditioning coordinate would have to solve (11) for all ρ simultaneously. But the vector fields $G^\rho(u)$, as ρ varies, span the full two-dimensional tangent space on the positive quadrant, so any common first integral is locally constant. Localization is therefore not a cosmetic choice; it is

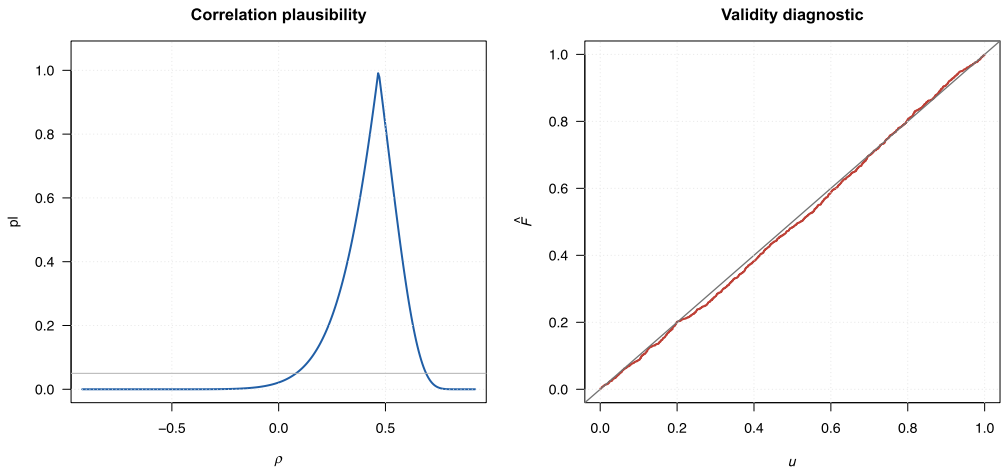


Figure 4. Bivariate normal correlation example. The left panel shows a local conditional plausibility curve for ρ when $n = 20$, $s_1 = 30.60$, and $s_2 = 10.52$. The right panel is a validity diagnostic based on 1200 simulations under $\rho = 0.45$; the empirical distribution of $\text{pl}_S(0.45)$ is close to the uniform benchmark.

the operation that leaves a one-dimensional auxiliary prediction problem while preserving conditional validity at the asserted value.

For a simulated data set with $n = 20$ generated at $\rho = 0.45$, the observed values are $s_1 = 30.60$ and $s_2 = 10.52$. Figure 4 shows the resulting local conditional plausibility curve and a validity diagnostic. Martin and Liu's simulations show that local conditional IM intervals attain the target coverage and can compare favorably with highly accurate conditional likelihood approximations based on $r^\star$ (Barndorff-Nielsen, 1986; Fraser, 1990; Martin & Liu, 2015a; Reid, 1995).

5. Marginal IMs

Scientific questions are usually marginal questions. Researchers ask about a treatment effect, a success probability, a variance component, a risk difference, or a prediction target, not about every unknown in a model simultaneously. Marginal IMs address inference on an interest parameter in the presence of nuisance parameters. The discussion below focuses on Martin and Liu (2015c).

5.1. Regular marginal IMs

Let $\theta = (\psi, \xi)$, where ψ is the interest parameter and ξ denotes the nuisance structure. A brute-force IM for the full θ may be valid, but it often predicts more auxiliary variation than the scientific question requires. This is inefficient, because predictive random sets become larger as the auxiliary dimension grows. The aim of a marginal IM is to construct a lower-dimensional association for ψ itself, then predict only the auxiliary component that remains genuinely uncertain for the targeted inference.

The regular case of Martin and Liu (2015c) makes the idea precise. Suppose a baseline association can be rewritten, after a smooth transformation of data and auxiliary variables,

in the form

$$\bar{p}(X, \psi) = \bar{a}(V_1, \psi), \quad c(X, V_2, \psi, \xi) = 0,$$

where $V = (V_1, V_2)$ has known distribution. The first equation involves the interest parameter and the auxiliary feature V_1; the second equation contains nuisance structure. The association is regular, in the sense of Martin and Liu (2015c), when the nuisance equation can be solved in ξ for any fixed (x, v_2, ψ) in the relevant range. Then V_2 carries no direct information about ψ. It can be ignored, or equivalently the predictive random set can be stretched to fill the entire V_2 direction, without damaging validity for ψ.

The resulting marginal association for inference about ψ is

$$\bar{p}(X, \psi) = \bar{a}(V_1, \psi), \quad V_1 \sim \mathsf{P}_{V_1}.$$

For observed x and auxiliary value v_1, define

$$\Psi_x(v_1) = \{\psi : \bar{p}(x, \psi) = \bar{a}(v_1, \psi)\}.$$

If $\mathcal{S}$ is a valid predictive random set for V_1, then

$$\Psi_x(\mathcal{S}) = \bigcup_{v_1 \in \mathcal{S}} \Psi_x(v_1)$$

induces marginal belief and plausibility functions

$$\text{mbel}_x(A) = \mathsf{P}_{\mathcal{S}}\{\Psi_x(\mathcal{S}) \subseteq A\}, \quad \text{mpl}_x(A) = \mathsf{P}_{\mathcal{S}}\{\Psi_x(\mathcal{S}) \cap A \neq \emptyset\}.$$

The validity theorem says, roughly, that valid prediction of V_1 gives valid marginal inference for ψ, uniformly over nuisance values ξ. This is the key point: nuisance elimination happens before probability calculus is transferred to the parameter space.

Like inference with basic IMs, it is often useful to summarize a marginal IM by a pointwise plausibility contour

$$\psi \mapsto \text{pl}_x(\psi). \tag{13}$$

Its level sets $\{\psi : \text{pl}_x(\psi) > \alpha\}$ give plausibility regions when the pointwise construction has the appropriate validity property. For more general assertions $A \subseteq \Psi$, a natural validity-preserving extension of the pointwise contour in (13) is

$$\text{pl}_x(A) = \sup_{\psi \in A} \text{pl}_x(\psi), \quad \text{bel}_x(A) = 1 - \sup_{\psi \notin A} \text{pl}_x(\psi).$$

5.2. Nonregular marginal IMs

Regularity is sufficient but not universal. In nonregular problems, the reduced auxiliary component relevant to ψ may still have a distribution depending on ξ. Then simply ignoring nuisance structure can lose validity. The marginal IM response is to use a uniformly valid predictive random set: a random set that predicts the nuisance-dependent auxiliary variable validly for every possible nuisance value. One practical route is to replace the nuisance-dependent auxiliary variable by a nuisance-free stochastic bound with fatter tails, then predict that bound. The resulting generalized marginal IM may be conservative, but the conservatism

is explicit and protects the plausibility scale. These are illustrated in detail below using the Behrens-Fisher example.

A differential-geometric formulation of nonregular marginal IMs, encompassing regular marginal IMs as special cases, remains an open problem with the potential to yield an elegant and unified mathematical theory.

5.3. Examples

5.3.1. A regular marginal IM example

The unknown-variance normal example is the cleanest illustration. If $\psi = \mu$ and $\xi = \sigma$, the conditional reduction to $(\bar{X}, S)$ gives

$$\bar{X} = \mu + \sigma n^{-1/2} U_1, \quad S = \sigma U_2,$$

where $U_1 \sim N(0,1)$ and $(n-1)U_2^2 \sim \chi^2_{n-1}$, independently. Dividing the centered mean equation by the observed scale removes σ and yields the marginal association

$$\frac{\bar{X} - \psi}{S/\sqrt{n}} = T, \quad T \sim t_{n-1}.$$

The plausibility contour for ψ follows directly from this one-dimensional auxiliary variable. This is what a good marginal procedure should do: answer the question about μ without requiring a prior distribution for σ and without first constructing a joint posterior-like object for (μ, σ). The success here should not be overgeneralized. It works because the nuisance scale can be removed by a pivot whose distribution is free of σ.

5.3.2. A nonregular marginal IM example

The Behrens–Fisher problem is the classical two-sample normal problem with unknown and unequal variances, originating in the work of Behrens (1929) and Fisher's fiducial analysis (Fisher, 1935). It has since served as a useful testing ground for fiducial, Bayesian, frequentist, and more recently IM approaches; see, for example, the practical discussion in Scheffé (1970), the review of Kim and Cohen (1998), the exact-solution comparison of Dudewicz et al. (2007), and the broader review by Paul et al. (2019). The marginal IM treatment in Martin and Liu (2015c) is especially relevant here because it shows how the example exposes the boundary between regular and nonregular marginalization. What is discussed below follows Wang and Liu (2026). Let

$$X_{ki} \overset{\text{ind}}{\sim} N(\mu_k, \sigma_k^2), \quad i = 1, \ldots, n_k, \; k = 1, 2,$$

with interest focused on

$$\psi = \mu_2 - \mu_1.$$

The nuisance structure consists of the remaining location direction and the two variances. This familiar example is useful here because regular marginalization almost works, but not quite.

After the conditional-IM reduction to sample means and sample standard deviations, the retained association is

$$\bar{X}_k = \mu_k + \sigma_k n_k^{-1/2} U_{1k}, \quad S_k = \sigma_k U_{2k},\ k = 1, 2,$$

where $U_{1k} \sim N(0,1)$ and $(n_k - 1)U_{2k}^2 \sim \chi^2_{n_k-1}$, independently. The reduction from the original four parameters to the two-parameter marginalization association proceeds in three steps. First, reparameterize

$$\lambda = \mu_1, \quad \psi = \mu_2 - \mu_1, \quad \omega = f(\sigma_1, \sigma_2), \quad \xi = \frac{\sigma_1^2/n_1}{\sigma_1^2/n_1 + \sigma_2^2/n_2},$$

where

$$f(a,b) = \left(\frac{a^2}{n_1} + \frac{b^2}{n_2}\right)^{1/2},$$

so that

$$\mu_1 = \lambda, \quad \mu_2 = \lambda + \psi, \quad \sigma_1 = \omega(n_1\xi)^{1/2}, \quad \sigma_2 = \omega\{n_2(1-\xi)\}^{1/2}.$$

Here λ is the irrelevant location level, ω is the overall standard-error scale, and $\xi \in (0,1)$ is the nuisance variance-ratio coordinate.

Second, subtract the two sample-mean associations. With $\bar{Y} = \bar{X}_2 - \bar{X}_1$,

$$\bar{Y} = \psi + \omega\{\sqrt{1-\xi}\, U_{12} - \sqrt{\xi}\, U_{11}\}.$$

For each fixed ξ, the auxiliary combination

$$U_1(\xi) = \sqrt{1-\xi}\, U_{12} - \sqrt{\xi}\, U_{11}$$

has a $N(0,1)$ distribution and is independent of (U_{21}, U_{22}). Thus, the nuisance location λ has disappeared, but the overall scale ω and the variance-ratio coordinate ξ remain.

Third, remove ω by studentizing with the observed standard errors. Since

$$\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} = \omega^2\{\xi U_{21}^2 + (1-\xi)U_{22}^2\},$$

one obtains

$$T_\psi(X) = \frac{\bar{X}_2 - \bar{X}_1 - \psi}{\{S_1^2/n_1 + S_2^2/n_2\}^{1/2}}.$$

This statistic satisfies

$$T_\psi(X) = \frac{U_1(\xi)}{\{\xi U_{21}^2 + (1-\xi)U_{22}^2\}^{1/2}}.$$

The remaining information about ξ comes from the observed ratio of sample standard errors:

$$R = \frac{n_1 S_2^2}{n_2 S_1^2} = \frac{1-\xi}{\xi}\frac{U_{22}^2}{U_{21}^2}.$$

Writing U_1 for a generic $N(0,1)$ variable with the same distribution as $U_1(\xi)$, the resulting two-parameter marginalization association for (ψ, ξ) is

$$T_\psi(X) = Z_1(\xi) = \frac{U_1}{\{\xi U_{21}^2 + (1-\xi)U_{22}^2\}^{1/2}}, \tag{14}$$

$$R = Z_2 \frac{1-\xi}{\xi}, \quad Z_2 = \frac{U_{22}^2}{U_{21}^2}. \tag{15}$$

The location λ and overall scale ω have been eliminated before any predictive random set is introduced. If the law of $Z_1(\xi)$ were free of ξ, this would be a regular marginal association for ψ. Instead, the law of $Z_1(\xi)$ changes with ξ: at the variance-ratio boundaries it approaches Student t_{n_1-1} or t_{n_2-1} limits, and in the balanced case the interior law is still different from the common boundary law. The second equation does not eliminate the nuisance ratio; solving it gives $\xi = Z_2/(Z_2 + R)$, which depends on the unobserved auxiliary Z_2. Thus, the exact reduced association remains genuinely two-dimensional.

Following the validity-first treatment in Wang and Liu (2026), which builds on the generalized marginal IM strategy of Martin and Liu (2015c), this nonregularity is handled by predicting a two-dimensional auxiliary variable, then using a projection that is sharp for inference on the marginal contrast. The key probability bound is due to Hsu (1938) and was later emphasized as a practical conservative solution by Scheffé (1970). It says that, for

$$m = \min(n_1, n_2) - 1,$$

the symmetric tails of $Z_1(\xi)$ are bounded uniformly by those of a Student T_m variable:

$$\mathsf{P}_\xi\{|Z_1(\xi)| > c\} \le \mathsf{P}\{|T_m| > c\}, \quad c \ge 0.$$

The bound is sharp at the least favorable variance-ratio boundary. The corresponding IM predictive random set is the cylinder

$$\mathcal{S}_{\mathrm{IM}} = [-C, C] \times \mathbb{R}_+, \quad C \sim |T_m|. \tag{16}$$

It is sharp in the mean-contrast coordinate and vacuous in the variance-ratio coordinate. The induced plausibility contour is

$$\mathrm{mpl}_x(\psi) = 2\left[1 - G_m\left(\left|\frac{\bar{x}_2 - \bar{x}_1 - \psi}{\{s_1^2/n_1 + s_2^2/n_2\}^{1/2}}\right|\right)\right], \tag{17}$$

where G_m is the cdf of T_m. The associated $100(1-\alpha)\%$ plausibility interval is the Hsu–Scheffé interval

$$\bar{x}_2 - \bar{x}_1 \pm t_{m,1-\alpha/2}\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^{1/2},$$

and it is valid uniformly over the unknown variances.

The important lesson is not only validity but the order in which validity and efficiency are judged. Since the nuisance ratio is not removed by regular marginalization, any prior-free fixed-sample method must either use a uniform bound or introduce additional structure. In the cylindrical class, Hsu's sharp boundary bound makes the IM interval minimax and admissible: no smaller two-sided critical value can retain uniform validity. More

generally, marginal inference on ψ depends only on the first-coordinate projection of a two-dimensional PRS, so attempts to predict the variance-ratio auxiliary can redistribute interval width across variance-ratio regimes but cannot uniformly shorten the projection without paying elsewhere.

This explains the relation with classical competitors. Welch's approximation (Welch, 1938, 1947), together with Satterthwaite's moment-matched degrees of freedom (Satterthwaite, 1946), gives the familiar Welch–Satterthwaite interval. It uses the observed standard errors to choose a data-dependent reference t distribution and is often shorter, but it is not an exact finite-sample procedure with uniform coverage over the variance ratio. Fisher's fiducial, generalized fiducial, and independent-Jeffreys Bayesian intervals use the observed ratio of standard errors in different ways; they can be reasonable, and in special comparisons even attractive, but their efficiency comparisons depend on the criterion used to weight variance-ratio regimes. The IM/Hsu answer is the validity-first answer: it is conservative because it refuses to average over or otherwise guess the nuisance variance ratio. The example therefore shows both the value and the limit of marginal IMs. Auxiliary variables can diagnose exactly where nuisance information remains, and validity can be preserved without a prior, but some nonregular problems force conservatism if the guarantee is to be uniform.

6. Conclusion

The main message is that scientific uncertainty should be both data-conditional and frequency-calibrated. The auxiliary-variable perspective connects these two demands when a genuine prior is not available. Inferential models continue a long historical project on inference with auxiliaries. Fisher developed fiducial inference as an attempt to obtain post-data probability statements about unknown parameters directly from the sampling model, by transferring the known randomness of pivotal or auxiliary quantities to the parameter space without introducing a prior distribution. Neyman supplied mathematical clarity through confidence procedures. Dempster and Shafer supplied a language of belief, plausibility, and lower and upper probabilities. Generalized fiducial inference revived the auxiliary-variable mechanism. IMs synthesize these ideas by treating the unobserved auxiliary variable as the carrier of inferential uncertainty and predicting it in a calibrated way, with efficiency pursued under the constraint of validity.

Several extensions and applications deserve further reading. Ryan Martin has played a leading role in advancing the IM framework, particularly its connections with imprecise probability and its possibilistic formulation, as reviewed in Martin (2026b). For imprecise-probability updating and the IM response to unsettling combination rules, see Gong and Meng (2021) and Liu and Martin (2021); for p-values as plausibilities, see Martin and Liu (2014); for the confidence-curve and fiducial reconstruction viewpoint, see Cui and Hannig (2025), Schweder and Hjort (2016) and Xie and Singh (2013). For modern generalized and possibilistic IMs based on calibrated contours, see Cahoon and Martin (2021), Jiang et al. (2024), Liu and Martin (2024) and Martin (2015, 2018, 2026a, 2026b); related betting-IM and e-process constructions are developed in Martin (2024, 2026c), Dixit and Martin (2025) and Dey et al. (2026). For non-parametric and high-dimensional directions, many-normal-means and shape-constrained examples provide useful entry points, including high-dimensional structural learning (Eschker & Liu, 2022) and partial conditioning for Hölder-constrained means (Yang et al., 2023); in these developments, explicit

reasoning with auxiliary variables introduced through the full data-generating association, as in the A/P/C construction described above, continues to play an important role.

For readers shaped by Jaynes' view of probability as the logic of science, IMs should not be read as a rejection of that logic (Jaynes, 2003). They sharpen its starting point. When the sampling model and a genuine prior distribution are both on the table, Bayes' theorem is the natural way to refine knowledge through the observed data. The Martin–Liu question is what the logic of science should do when the prior entry is empty. The IM answer is not automatically to fill that empty place with a formal or default prior and then interpret the resulting posterior as objective scientific uncertainty. It is to keep on the table what is still genuinely available: the sampling model, the auxiliary variable, its known distribution, the observed association, and the scientific assertion, some of which become redundant and, thereby, often hidden when a prior is available. In this sense, Jaynes teaches that probability is the right logic when the information warrants a probability distribution; Martin and Liu teach that the logic of science must also say what to do when it does not. Thus, the development of IMs using auxiliary variables can be read as continuing Jaynes' project of probability as the logic of science, but with an important qualification: when the available information does not justify a full prior probability distribution, the logic of uncertainty should be calibrated, auxiliary-variable based, and sometimes imprecise.

Nevertheless, to invite further reflection, Bayesian prior specification may itself be viewed as a further stage of model building, even when genuine prior knowledge about the scientifically meaningful unknowns already introduced in the model is unavailable; see Box (1980), Gelman and Shalizi (2013) and Rubin (1984). This perspective raises questions about the interpretation and justification of such specifications, which we leave for deeper discussion elsewhere. Ongoing research also explores how the auxiliary-variable perspective can clarify the likelihood principle. For IMs, a central question is whether a specified calibrated plausibility contour can be recovered from likelihood alone, or whether additional information about the auxiliary reference law and conditioning structure is essential. Likelihood compares parameter values through the observed data, while calibration draws on the probability law governing the relevant auxiliary uncertainty. Understanding when likelihood retains the structure needed for this calibration, and what additional information is required when it does not, provides a concrete direction for foundational investigation.

There is still much to do beyond expanding the range of IM applications. For continuous-data models, a differential-geometric theory of IMs now seems within reach: conditioning, marginalization, auxiliary reductions, and local validity all have natural geometric interpretations in terms of transformations, fibers, transverse coordinates, and induced contours. Discrete-data models are more challenging because the clean geometry of smooth auxiliary maps is replaced by jumps, nonunique inversions, and boundary-sensitive validity.

Acknowledgments

The author thanks the editors and reviewers for their helpful and constructive comments, which have improved the clarity and readability of this article.

Author contributions

CRediT: **Chuanhai Liu:** Writing – original draft

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported in part by the U.S. National Science Foundation [grant number DMS-2412629].

ORCID

Chuanhai Liu http://orcid.org/0000-0003-1919-7280

References

Balch, M. S., Martin, R., & Ferson, S. (2019). Satellite conjunction analysis and the false confidence theorem. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *475*(2227), 20180565.

Barnard, G. A. (1995). Pivotal models and the fiducial argument. *International Statistical Review*, *63*(3), 309–323. https://doi.org/10.2307/1403482

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, *73*(2), 307–322. https://doi.org/10.1093/biomet/73.2.307

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Behrens, W. U. (1929). Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen. *Landwirtschaftliche Jahrbücher*, *68*, 807–837.

Berger, J. O., Bernardo J. M., & Sun, D. (2009). The formal definition of reference priors. *The Annals of Statistics*, *37*(2), 905–938. https://doi.org/10.1214/07-AOS587

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society: Series B*, *41*(2), 113–147. https://doi.org/10.1111/j.2517-6161.1979.tb01066.x

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society, Series A (General)*, *143*(4), 383–404.

Cahoon, J., & Martin, R. (2021). Generalized inferential models for censored data. *International Journal of Approximate Reasoning*, *137*, 51–66. https://doi.org/10.1016/j.ijar.2021.06.015

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*(4), 404–413. https://doi.org/10.1093/biomet/26.4.404

Cui, Y., & Hannig, J. (2025). Demystifying inferential models and confidence curves: A fiducial perspective. *Statistical Science*, *40*(2), 211–218. https://doi.org/10.1214/24-STS924

Dempster, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *The Annals of Mathematical Statistics*, *37*(2), 355–374. https://doi.org/10.1214/aoms/1177699517

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, *38*(2), 325–339. https://doi.org/10.1214/aoms/1177698950

Dempster, A. P. (1968). Upper and lower probabilities generated by a random closed interval. *The Annals of Mathematical Statistics*, *39*(3), 957–966. https://doi.org/10.1214/aoms/1177698328

Dempster, A. P. (2008). The Dempster–Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, *48*(2), 365–377. https://doi.org/10.1016/j.ijar.2007.03.004

Dey, N., Martin, R., & Williams, J. P. (2026). Generalized universal inference on risk minimizers. *Journal of the Royal Statistical Society: Series B*, *88*(3), 756–778. https://doi.org/10.1093/jrsssb/qkaf065

Dixit, V., & Martin, R. (2025). Anytime-valid and asymptotically efficient inference driven by predictive recursion. *Biometrika*, *112*(2), asae066. https://doi.org/10.1093/biomet/asae066

Dudewicz, E. J., Ma, Y., Mai, E. S., & Su, H. (2007). Exact solutions to the Behrens–Fisher problem: Asymptotically optimal and finite sample efficient choice among. *Journal of Statistical Planning and Inference*, *137*(5), 1584–1605. https://doi.org/10.1016/j.jspi.2006.09.007

Ermini Leaf, D., & Liu, C. (2012). Inference about constrained parameters using the elastic belief method. *International Journal of Approximate Reasoning*, *53*(5), 709–727. https://doi.org/10.1016/j.ijar.2012.02.003

Eschker, S. J., & Liu, C. (2022). Estimation of high-dimensional normal means through inferential models. arXiv:2207.05190.

Fisher, R. A. (1935). The fiducial argument in statistical inference. *Annals of Eugenics*, *6*(4), 391–398. https://doi.org/10.1111/j.1469-1809.1935.tb02120.x

Fisher, R. A. (1973). *Statistical Methods and Scientific Inference* (3rd ed.). Hafner Press.

Fraser, D. A. S. (1968). *The Structure of Inference*. John Wiley & Sons.

Fraser, D. A. S. (1990). Tail probabilities from observed likelihoods. *Biometrika*, *77*(1), 65–76. https://doi.org/10.1093/biomet/77.1.65

Fraser, D. A. S. (2004). Ancillaries and conditional inference. *Statistical Science*, *19*(2), 333–369. https://doi.org/10.1214/088342304000000323

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

Fraser, D. A. S., Bédard, M., Wong, A., Lin, W., & Fraser, A. M. (2016). Bayes, reproducibility and the quest for truth. *Statistical Science*, *31*(4), 578–590.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press.

Gelman, A., & Shalizi, C. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 8–38.

Ghosh, M., Reid, N., & Fraser, D. A. S. (2010). Ancillary statistics: A review. *Statistica Sinica*, *20*(4), 1309–1332.

Gong, R., & Meng, X. L. (2021). Judicious judgment meets unsettling updating: Dilation, sure loss and Simpson's paradox. *Statistical Science*, *36*(2), 169–190. https://doi.org/10.1214/19-STS765

Hannig, J. (2009). On generalized fiducial inference. *Statistica Sinica*, *19*(2), 491–544.

Hannig, J., Iyer, H., Lai, R. C. S., & Lee, T. C. M. (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, *111*(515), 1346–1361. https://doi.org/10.1080/01621459.2016.1165102

Hsu, P. L. (1938). Contributions to the theory of 'Student's' *t*-test as applied to the problem of two samples. In *Statistical Research Memoirs* (Vol. 2, pp. 1–24). University College.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620–630. https://doi.org/10.1103/PhysRev.106.620

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London Series A*, *186*(1007), 453–461. https://doi.org/10.1098/rspa.1946.0056

Jeffreys, H. (1948). *Theory of Probability* (2nd ed.). Oxford University Press.

Jiang, Y., Liu, C., & Zhang, H. (2024). Finite sample valid inference via calibrated bootstrap. arXiv:2408.16763.

Kim, S. H., & Cohen, A. S. (1998). On the Behrens–Fisher problem: A review. *Journal of Educational and Behavioral Statistics*, *23*(4), 356–377. https://doi.org/10.3102/10769986023004356

Liu, C. (2026). *Differential Geometry of Conditional Inferential Models: Obstructions, Reassociation, and Connections to Fisher Information*. Unpublished manuscript, Department of Statistics, Purdue University.

Liu, C., & Martin, R. (2015). Frameworks for prior-free posterior probabilistic inference. *WIREs Computational Statistics*, *7*(1), 77–85. https://doi.org/10.1002/wics.1329

Liu, C., & Martin, R. (2021). Comment: Settle the unsettling: An inferential models perspective. *Statistical Science*, *36*(2), 196–200. https://doi.org/10.1214/21-STS765B

Liu, C., & Martin, R. (2024). Inferential models and possibility measures. In Berger J. O., Meng X. L. , Reid N., & Xie M. G. (Eds.), *Handbook of Bayesian, Fiducial, and Frequentist Inference* (pp. 344–363). Chapman & Hall/CRC. https://doi.org/10.1201/9780429341731

Martin, R. (2015). Plausibility functions and exact frequentist inference. *Journal of the American Statistical Association*, *110*(512), 1552–1561. https://doi.org/10.1080/01621459.2014.983232

Martin, R. (2018). On an inferential model construction using generalized associations. *Journal of Statistical Planning and Inference*, *195*, 105–115. https://doi.org/10.1016/j.jspi.2016.11.006
Martin, R. (2019). False confidence, non-additive beliefs, and valid statistical inference. *International Journal of Approximate Reasoning*, *113*, 39–73. https://doi.org/10.1016/j.ijar.2019.06.005
Martin, R. (2024). Ryan Martin's contribution to the discussion of 'estimating means of bounded random variables by betting' by Waudby-Smith and Ramdas. *Journal of the Royal Statistical Society: Series B*, *86*(1), 43–44. https://doi.org/10.1093/jrsssb/qkad112
Martin, R. (2026a). An efficient Monte Carlo method for valid prior-free possibilistic statistical inference. *Journal of the American Statistical Association*, 121, 1–25. https://doi.org/10.1080/01621459.2026.2671450
Martin, R. (2026b). Possibilistic inferential models: A review. *Journal of the American Statistical Association*, *121*(553), 807–826. https://doi.org/10.1080/01621459.2025.2606127
Martin, R. (2026c). Regularized E-processes: Anytime valid inference with knowledge-based efficiency gains. *Bernoulli*, *32*(3), 1947–1974. https://doi.org/10.3150/25-BEJ1939
Martin, R., & Liu, C. (2013). Inferential models: A framework for prior-free posterior probabilistic inference. *Journal of the American Statistical Association*, *108*(501), 301–313. https://doi.org/10.1080/01621459.2012.747960
Martin, R., & Liu, C. (2014). A note on *p*-values interpreted as plausibilities. *Statistica Sinica*, *24*(4), 1703–1716. https://doi.org/10.5705/ss.2013.087
Martin, R., & Liu, C. (2015a). Conditional inferential models: Combining information for prior-free probabilistic inference. *Journal of the Royal Statistical Society: Series B*, *77*(1), 195–217. https://doi.org/10.1111/rssb.12070
Martin, R., & Liu, C. (2015b). *Inferential Models: Reasoning with Uncertainty*. Chapman and Hall/CRC. https://doi.org/10.1201/b19269
Martin, R., & Liu, C. (2015c). Marginal inferential models: Prior-free probabilistic inference on interest parameters. *Journal of the American Statistical Association*, *110*(512), 1621–1631. https://doi.org/10.1080/01621459.2014.985827
Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London Series A*, *236*(767), 333–380. https://doi.org/10.1098/rsta.1937.0005
Neyman, J. (1941). Fiducial argument and the theory of confidence intervals. *Biometrika*, *32*(2), 128–150. https://doi.org/10.1093/biomet/32.2.128
Paul, S. R., Wang, Y. G., & Ullah, I. (2019). A review of the Behrens–Fisher problem and some of its analogs: Does the same size fit all? *REVSTAT–Statistical Journal*, *17*(4), 563–597. https://doi.org/10.57805/revstat.v17i4.281
Reid, N. (1995). The roles of conditioning in inference. *Statistical Science*, *10*(2), 138–157. https://doi.org/10.1214/ss/1177010027
Rubin, D. B. (1981). Estimation in parallel randomized experiments. *Journal of Educational Statistics*, *6*(4), 377–401. https://doi.org/10.3102/10769986006004377
Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, *12*(4), 1151–1172. https://doi.org/10.1214/aos/1176346785
Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110–114. https://doi.org/10.2307/3002019
Scheffé, H. (1970). Practical solutions of the Behrens–Fisher problem. *Journal of the American Statistical Association*, *65*(332), 1501–1508. https://doi.org/10.1080/01621459.1970.10481179
Schweder, T., & Hjort, N. L. (2016). *Confidence, Likelihood, Probability: Statistical Inference with Confidence Distributions*. Cambridge University Press. https://doi.org/10.1017/CBO9781139046671
Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.
Wang, X., & Liu, C. (2026). Revisiting the Behrens–Fisher problem: Validity-first optimality. arXiv:2606.07847. https://doi.org/10.48550/arXiv.2606.07847
Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, *29*(3/4), 350–362. https://doi.org/10.1093/biomet/29.3-4.350
Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, *34*(1/2), 28–35. https://doi.org/10.1093/biomet/34.1-2.28

Xie, M. G., & Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review*, *81*(1), 3–39. https://doi.org/10.1111/insr.12000
Yang, J., Wang, X., & Liu, C. (2023). Partial conditioning for inference of many-normal-means with Hölder constraints. *International Journal of Approximate Reasoning*, *159*, 108946. https://doi.org/10.1016/j.ijar.2023.108946
Zabell, S. L. (1992). R. A. Fisher and the fiducial argument. *Statistical Science*, *7*(3), 369–387.