3

Distributions for Parameters

N. Reid
University of Toronto

3.1 Introduction

This handbook is the outcome of a series of workshops, as described in the preface. In this introductory chapter, I attempt to give a high-level overview of different inferential approaches. I have addressed this through what I view as a common goal among Bayesian, fiducial and frequentist approaches — finding a distribution for a parameter or parameters of interest. Confidence distributions, generalized fiducial distributions, inferential models, and belief functions are some of the terms associated with these approaches. In this overview, I discuss both historical developments and modern versions. For a more detailed comparison of Bayesian and frequentist approaches, see Berger (2022). A more eclectic discussion of Bayes, Fiducial and Frequentist is given in Meng (2024) though a series of compelling examples. An exploration of frequentist properties of Bayesian methods in a quite different context, of testing fixed hypotheses, is presented in Berger *et al.* (2024).

The history of statistical inference, while young, is surprisingly rancorous. The hostility between Fisher and Neyman may have contributed to a divergence of developments in the United Kingdom and in the United States that hampered the maturing of the subject (Zabell, 1992; Lehmann, 1993). About Bayesian inference, Fisher (1930) wrote:

> I know of only one case in mathematics of a doctrine which has been accepted and developed by the most eminent men of their time, and is now perhaps accepted by men now living, which at the same time has seemed to a succession of sound writers to be fundamentally false and devoid of foundation.

The discussions at the BFF workshops are much more congenial. One theme of various approaches discussed there is the desire to obtain a distribution for parameters of a statistical model that enables us to make inferential statements. In this chapter, I give an introduction to BFF from this point of view. The topics are treated in more depth in other chapters in this volume and, true to our history, present divergent viewpoints.

3.2 An Historical Sketch

The most familiar way to describe a distribution for parameters is to use the calculus of conditional probability embodied in Bayes' theorem. We assume a model for the observation of a random process, described by a family of probability density functions for possible

DOI: 10.1201/9780429341731-3

outcomes, denoted $\{f(y;\theta), \theta \in \Theta\}$; the form of the density function is assumed known. We further assume that y takes values in $\mathcal{Y}$ and that the sets $\mathcal{Y}$ and Θ are known. Given a probability distribution for θ with density function $\pi(\theta)$, and interpreting $f(y;\theta)$ as the conditional density for y, given θ, standard probability calculus gives the conditional density for θ, given y:

$$\pi(\theta \mid y) = \frac{f(y;\theta)\pi(\theta)}{\int f(y;\theta)\pi(\theta)d\theta} = \frac{f(y;\theta)\pi(\theta)}{m(y)}. \tag{3.1}$$

We assume that $m(y)$, the marginal density of y, is positive in this definition; a more rigorous formulation is given in, for example, the introductory chapters of Schervish (1995). In statistical applications y is the data, often a sample of size n, $\pi(\theta)$ is called the prior density for θ, and $\pi(\theta \mid y)$ is called the posterior density. The informal notion is that $\pi(\theta)$ is available to the experimenter before any data are collected, and the data modify the prior to give a probability density for θ suitable for inference once y has been observed. Stigler (2013) noted that the original title of Bayes' paper was "A method of calculating the exact probability of all conclusions based on induction", but the published title was somewhat more tentative (Bayes, 1763).

The phrase "all conclusions" is widely relied upon in modern applications of Bayesian inference. Armed with the posterior density, the statistician can compute probabilities of any events of interest, such as $\Pr(\theta > 0 \mid y)$, or $\Pr(\theta_1 > \theta_2 \mid y)$. Thus a potential conclusion from a Bayesian analysis of data from an experiment to compare treatments A and B might take the form "from this data and model we see that the (posterior) probability that μ_A is larger than μ_B is 0.09"; μ_A and μ_B here being parameters in the statistical model for the responses. Kline *et al.* (2021) reported on a Bayesian analysis of a household survey in Ohio, undertaken in 2020 to assess the sero-prevalence of antibodies to the SARS-COV-2 virus. One of their conclusions is

> Based on our model, the posterior mean prevalence is 1.3% with a 95% credible interval of (0.2%, 2.7%).

The data are fixed in this conclusion, so the probability referred to essentially comes from the prior distribution. As a result, much turns on the choice of the prior and its meaning. In Bayes (1763), the parameter space Θ is $[0, 1]$ and $\pi(\theta)$ is a uniform density on Θ, a choice that was supported by Laplace; see Stigler (1986). It was quickly realized that a uniform prior is not invariant to the parametrization of the model, so many attempts were made to find priors by other means. Jeffreys (1961) used parametrization invariance to motivate priors based on Fisher information; see Section 3.4.1 and Kass and Wasserman (1996, Section 2).

In subjective Bayesian inference (Savage, 1954), emphasis is placed on formulating a proper prior distribution from consideration of probabilities of events about θ in the absence of data. This can be useful as an aid to personal decision-making, but is less helpful for establishing a scientific consensus.

Often inference is labelled "Bayesian" in settings where there is considerable empirical evidence for the distribution of the parameters in a model. For example, in Kline *et al.* (2021), the model parameters include some components related to the sensitivity and specificity of the antibody tests, information that was externally available from manufacturers as part of their validation studies. Strictly speaking, the results then follow directly from probability calculus, and it is debatable whether this should be described as a Bayesian analysis. In this chapter, and in most discussions of BFF, we consider that such concrete prior information is not available.

The fiducial approach to inference proposed in Fisher (1930) was an attempt to deduce a probability distribution from the data without using a prior distribution. Although Cox

(2016) described this as "almost certainly a major error", there have been many attempts to both make Fisher's ideas more precise and to put fiducial inference on a more solid footing. Dawid (2024) and Lang (2024, Section 15.4) give more detailed summaries of Fisher's original development. Taraldsen and Lindqvist (2024) make a distinction between the fiducial model and the statistical model. Vovk (2024, Section 17.2) discusses Fisher's approach to fiducial prediction.

Suppose there is some one-dimensional summary of the data, which Fisher denoted by T, and a one-dimensional parameter of interest, θ. As a very simple example, suppose the model for T is a location model with known density,

$$T = \theta + e, \quad e \sim f_0. \tag{3.2}$$

Once $T = t$ is observed, we can write $\theta = t - e$, and can think of translating to θ the probability model assigned to the error variable, e. More generally, if we fix a percentile of the distribution function for T, $F(t;\theta)$, we create a link between T and θ. Fisher (1930) expressed this as

$$\mathrm{d}f = -\frac{\partial}{\partial\theta}F(t;\theta)\mathrm{d}\theta \tag{3.3}$$

and called this the fiducial probability density function for θ. The location model explication of fiducial inference is due to Fraser (1966), who called it a structural density, to distinguish it from Fisher's construction. The structural density was defined for models based on groups, as the link between observation and parameter is direct in such models. Fisher (1930) wrote "this then is a definite probability statement about the unknown parameter". Fraser (1968, p. 20) states "A probability statement concerning e is *ipso facto* a probability statement concerning θ".

Confidence distributions, described in Cox (1958), are closely related to fiducial distributions, but are perhaps more accessible. Neyman's original theory (Neyman, 1937) described the construction of an interval of parameter values at a given confidence level, often taken to be 0.95. The interpretation of the confidence level is indirect: in repeated sampling from the model, 95% of such confidence intervals will include the true value of the parameter. This is a statement about the statistical procedure, not the particular interval that has just been constructed; it is like a warranty from the statistician. Berger (2022) calls this "procedural frequentism". Cox (1958) suggested:

> ... in simple cases ... there seems no reason why we should not work with confidence distributions for the unknown parameter ... These can either be defined directly, or ... introduced in terms of the set of all confidence intervals.

While confidence intervals of the form $[a, b]$ are nearly always used in practice, it is easier to define a confidence distribution using a set of one-sided intervals of the form $(-\infty, a]$, following Efron (1993):

> Let $\theta_x(\alpha)$ be the upper endpoint of an exact or approximate one-sided level-α confidence interval for θ. ... The confidence distribution for θ is defined to be the distribution having density
>
> $$\pi_x^\dagger(\theta) = \mathrm{d}\alpha_x(\theta)/\mathrm{d}\theta; \tag{3.4}$$

here $\alpha_x(\theta)$ is the inverse function to $\theta_x(\alpha)$. Efron added:

> ... assigns probability 0.05 to θ lying between the upper endpoints of the 0.90 and 0.95 confidence intervals, etc. ... Of course this is logically incorrect, but it has powerful intuitive appeal. ... *in the case of no nuisance parameters, [this] is exactly Fisher's fiducial distribution* [emphasis added].

Confidence distributions were called significance functions in Fraser (1991), to reflect his construction based on the computation of p-values:

> ... significance [p-value] records the probability left of the observed data point... the significance function is a plot of this probability as a function of θ ... the full spectrum of confidence intervals is obtained ... suggesting the alternate name confidence distribution function.

I added the emphasis in the quote from Efron (1993) because accommodating nuisance parameters is the rock on which all these constructions have foundered, and one reason there are active investigations and extensions underway today. A distribution function is most useful if it is available for a single parameter of interest, but statistical models useful for the analysis of data have many unknown parameters. Even if our data are assumed to come from a single normal distribution, we may in different circumstances be interested in the expected value of that distribution, with the variance parameter treated as a "nuisance", or vice versa. In a fully Bayesian setting, we can in principle construct a joint posterior distribution for all the parameters in the model, and then integrate over the nuisance parameters to obtain a marginal posterior distribution for the parameter of interest. It is not possible to apply the same reasoning to fiducial distributions, confidence distributions, or significance functions, because the probability that underlies their interpretation is fundamentally different.

Marginalization of probabilities for vector parameters to probabilities for scalar parameters is one version of the combination of probabilities, but the probability calculus for computing the combined probabilities of several events also does not apply to fiducial and confidence distributions. This is addressed in Dawid (2024, Section 5.6) and in Taraldsen and Lindqvist (2024). Example 5.8 in Dawid (2024), originally due to Stein (1959), is a simple but informative illustration of this. As a result, recent work on fiducial, confidence and significance functions has needed to address the combination problem separately. In Section 3.4, I will look at the modern approaches being developed as part of the BFF discussions. Before turning to this, it is helpful to consider the meaning of probability.

3.3 The Nature of Probability

Reid and Cox (2015), following many other writers on the subject, including Fisher (1956, p. 31), describe two meanings of probability. One meaning, sometimes called *epistemic* probability, describes uncertainty of knowledge. This is most often associated with Bayesian approaches, and the uncertainty may be, as Jeffreys (1961) proposed, a rational or impersonal degree of belief, or, as in Savage (1954), a personal degree of belief. Personal degrees of belief are sometimes described through a willingness to bet with different odds on the outcome of an event, and underpin subjective Bayesian inference.

The second meaning, usually called *empirical* or *aleatory* probability, encapsulates properties of a "real" world, either through the postulated existence of an infinite population, or a process of repetition under constant conditions. These will of necessity be hypothetical or highly idealized, but may still be suitable for the development and reporting of conclusions based on data. A textbook example of aleatory probability is the meaning of the statement "the probability of tossing heads with a fair coin is 0.5"; usually explained to students as a convergence to 0.5 of the observed fraction of heads in an endlessly repeated sequence. Berger (2022) calls this "empirical frequentism". Lang (2024, Section 15.3), gives a different view on the meaning of probability; for example, he describes both epistemic and aleatory probabilities as subjective; the subjectivity in the latter coming from the uncertain knowledge of the modeller.

It is not always very clear in the literature which meaning of probability is intended to be ascribed to fiducial, structural, or confidence, distributions, but in general it seems to be aleatory probability. That is, we expect that an inferential interval for a parameter θ which is ascribed a probability of 95%, should, in hypothetical repeated sampling from the model, include the true value of the parameter 95% of the time. Cox (1958) introduced the confidence distribution as simply a concise way to present a continuum of such (nested) intervals, instead of specifying a single level of confidence. On the other hand, in their book, Schweder and Hjort (2016, Chapter 1.6) refer to "objective epistemic probability". As I understand their discussion, they suggest that the interpretation of the confidence distribution is epistemic, but its properties in repeated sampling can be verified empirically.

In contrast, the probability that is ascribed to a Bayesian statement, based on a posterior distribution, is an epistemic probability describing the uncertainty in our knowledge. This uncertainty is derived from our uncertainty about the value of the parameters that govern the model, i.e., from the prior distribution. Taraldsen and Lindqvist (2024, Section 18.1) state that the interpretation of fiducial probability "is exactly as for a Bayesian posterior", which suggests an epistemic interpretation.

The calculus of probability, following axioms set out by Kolmogorov, is indifferent to the meaning ascribed to probabilistic statements. This calculus specifies in particular, that marginal probabilities can be obtained by summing or integrating joint probabilities, and that the total probability of all possible outcomes is 1, so that the probability of a set in the sample space is one minus the probability of its complement, and so on. Lang (2024) state that only aleatory probability satisfies Kolmogorov's axioms; Schweder and Hjort (2016, p. 7) write "No axiomatic theory has been worked out for epistemic probability to be used in science, except for Bayesian probabilities".

It is proposed in Reid and Cox (2015, Section 2.2) that by requiring Bayesian probability statements to have a calibration property in repeated sampling from the model, we avoid the need for two potential interpretations of probability. In particular, if

$$\Pr\{\theta \geq \theta^{(1-\alpha)}(y) \mid y\} = 1 - \alpha \tag{3.5}$$

then we should require as well that

$$\Pr\{\theta \geq \theta^{(1-\alpha)}(Y) \mid \theta\} = 1 - \alpha, \tag{3.6}$$

where the first probability statement comes from a Bayesian posterior distribution and is thus conditional on the observed data, and the second comes from the statistical model, and is thus conditional on the parameter. Berger (2022) argued that "even [subjective] Bayesians should accept the empirical frequentist principle"; see also Berger (2006).

If it were possible to choose a single prior distribution that ensures this calibration, then the discussion of inferential approaches would have ended many years ago. However, the current position is that this is not possible, using the notions of probability described so far.

Approaches to inference using an expanded notion of probability, such as Dempster–Shafer belief functions (Gong, 2024), or imprecise probabilities (Walley, 1990), have been argued to be the only way to ensure such calibration. While important for the foundations, these methods have not yet found widespread application, perhaps because the conclusions are more difficult to understand, and thus to explain to non-experts.

The search for distributions for parameters has re-emerged in recent literature, with a view to extending the ideas beyond the fairly limited range of models treated in the past. In the next section, I describe some of the progress that has been made, and the gaps that remain.

3.4 Recent Developments

3.4.1 Objective Bayes

"Objective Bayes" generally means any method proposed for deriving a posterior distribution that is not based on personal probability or subjective prior distributions. Classes of objective priors include the reference priors of Berger and Bernardo (1992) and probability matching priors, reviewed in Datta and Mukerjee (2004), but there are many others. The fact that so many variants exist suggests there is a problem in making the notion of objectivity precise. Kass and Wasserman (1996) provided an excellent overview of various approaches to what they call the formal selection of priors.

As discussed in Section 3.3, one way to avoid using epistemic probability in Bayesian inference is to consider this as one inferential method among many, and test it against its properties in sampling from the model. Thus, the probability referred to in the posterior conclusion becomes an empirical probability about the method. Reid and Cox (2015) write

> ... as with other measuring devices, within this scheme of repetition probability is defined as a hypothetical frequency... it is unacceptable if a procedure yielding high-probability regions in some non-frequency sense are poorly calibrated ... such procedures, used repeatedly, give misleading conclusions.

This suggests that for a given prior distribution on θ, if we compute a posterior quantile, say $\theta_\pi^{(1-\alpha)}(y)$, as at (3.5), now with the dependence on the prior explicitly noted, then we need to investigate the distribution of $\theta_\pi^{(1-\alpha)}(Y)$ under the model $f(y;\theta)$ for the observed random variable. It turns out that if $\theta \in \mathbb{R}$ is a scalar parameter, then Jeffreys' prior provides this type of calibration, at least approximately. Writing $\ell(\theta; y) = \log f(y;\theta)$ for the log-likelihood function of the model, the expected Fisher information is

$$i(\theta) = \mathrm{E}\{-\partial^2 \ell(\theta; Y)/\partial\theta^2\}. \tag{3.7}$$

The prior density

$$\pi_J(\theta) \propto i^{1/2}(\theta) \tag{3.8}$$

is the unique prior for which

$$\Pr\{\theta_\pi^{(1-\alpha)}(Y) \le \theta \mid \theta\} = 1 - \alpha + O(n^{-1}), \tag{3.9}$$

when $Y = (Y_1, \ldots, Y_n)$ is an independent sample of size n from the model $f(y;\theta)$. In (3.9) the term $O(n^{-1})$ is a remainder that roughly speaking looks like A/n, where $A < \infty$ may depend on θ, but not on n.

It is necessary to consider a remainder term of $O(n^{-1})$ in equation (3.9), as any proper prior will lead to posterior limits that are calibrated to $O(n^{-1/2})$; sometimes described as the "prior is washed out by the data". This follows from the central limit theorem, usually in this context called the Bernstein-von Mises theorem, for the posterior distribution. This first-order equivalence prompted the title of Fraser (2011). Kleijn and van der Vaart (2012) is a helpful reference on the Bernstein-von Mises theorem, comparing the usual case to that under model misspecification.

The practical appeal of Bayesian approaches is in multi-parameter problems, where we may write, for example,

$$\pi(\psi \mid y) = \int_{\{\theta:\psi(\theta)=\psi\}} \pi(\theta \mid y) d\theta. \tag{3.10}$$

If $\theta \in \mathbb{R}^d$, the expected Fisher information is a $d \times d$ matrix, and (3.8) becomes

$$\pi_J(\theta) \propto |i(\theta)|^{1/2}, \tag{3.11}$$

based on the determinant of the information matrix. Unfortunately, except in very special settings, posterior probability bounds computed using this prior, and marginalized as at (3.10), are not calibrated, even approximately:

$$\Pr\{\psi^{(1-\alpha)}(Y) \leq \psi \mid \theta\} \neq 1 - \alpha + O(n^{-1}). \tag{3.12}$$

There is nothing special about Jeffreys' prior in this disappointment: there is rarely a single prior for a vector parameter that provides calibration for arbitrary derived scalar parameters.

An illuminating and well-studied example due to Stein (1959) is the "many-normal-means" problem: $y_i \sim N(\theta_i, 1/n), i = 1, \ldots, d$, with the parameter of interest $\psi = \Sigma_i \theta_i^2$. Jeffreys' prior for θ is proportional to 1, and the posterior for θ, given y, is multivariate normal with expected value y and covariance matrix I/n, where I is the $d \times d$ identity matrix. The posterior density for ψ is a non-central χ_d^2, and posterior probability bounds based on this diverge from their nominal value in the sampling model quite dramatically. Stein (1959) showed, in the context of fiducial inference, that the expected value under the model of the mean of the marginal posterior density is $\psi + 2d/n$, whereas the expected value of the maximum likelihood estimator is $\psi + d/n$, and emphasized the divergence of approximate confidence intervals that follows from that. The divergence of the entire marginal posterior density as d increases with n fixed was noted in Fraser and Reid (1992) and studied in more detail in Reid and Sun (2010).

That flat priors are not well-calibrated for components or scalar functions of θ is a version of the marginalization paradox of Dawid *et al.* (1973); see also Dawid (2024, Sections 5.6 and 5.7.2), but was also noted in earlier literature in the context of particular examples. It was re-discovered in the particle physics literature in the 1990s, when many simulation experiments were being developed in preparation for data from the Large Hadron Collider (Heinrich, 2006). In spite of this lack of calibration, there are many applications of Bayesian inference in complex models where the authors say something along the lines of "we used standard vague priors for the parameters of our model", also assume independence of these priors, and proceed without attention to the properties of the resulting inference. There are somewhat fewer applied papers where checks on the stability of the inference to the form of the prior are included as part of the model-checking. There is an important theoretical literature on robustness of Bayesian inference (e.g., Berger *et al.*, 2000) and on prior-data conflict (e.g., Evans and Moshonov, 2006).

A "flat" prior for an unbounded parameter, while not a proper probability density function, can often lead to a proper posterior density, i.e., a density that is finite and integrates to 1. Many objective priors are defined this way, including Jeffreys' prior above, the uniform prior $\pi_U(\theta) \propto 1$, and most matching and reference priors. There are examples in the literature, however, where improper priors lead to improper posterior densities. If these densities are estimated by simulation schemes such as Markov chain Monte Carlo, the results are meaningless as the algorithm cannot converge.

Since it is not possible in general to obtain calibrated posterior distributions for arbitrary parameters of interest from a single prior, objective priors leading to calibration need to be targetted on the parameter of interest. This is explicitly recognized in the work on reference priors (Berger and Bernardo, 1992), and matching priors, but the theoretical results have not percolated very widely into practice. Gelman *et al.* (2008) suggested the use of weakly informative priors, with the expressed goal of finding a "somewhat informative prior distribution that can nonetheless be used in a wide range of applications". The weakly informative prior distribution is chosen to be proper and to put positive probability on all "plausible" values of the unknown parameter or parameters. The general calibration properties of the posteriors from weakly informative priors seems not to have been studied; numerical work for models widely used in ecology is provided in Lemoine (2019). If such priors are meant to give calibrated inferences, this would need to be checked in each instance, and, as usual, for each parameter of interest.

Another feature of priors leading to calibrated posterior inference for a parameter is that the priors necessarily depend on the form of the model, for example, Jeffreys' prior depends on the expected Fisher information. As another example, the probability matching priors developed by explicitly requiring (3.6) to hold to $O(n^{-1})$ (Peers, 1965; Mukerjee and Ghosh, 1997) must satisfy an equation of the form

$$\sum_{a=1}^{d} \frac{\partial}{\partial \theta_a}[\{i^{11}(\theta)\}^{-1/2} i^{a1}(\theta)\pi(\theta)] = 0, \tag{3.13}$$

where $i^{ab}(\theta)$ is the (a, b) entry of $i^{-1}(\theta)$ and we assume that the first component of θ is the parameter of interest. If the Fisher information matrix has $i_{1a}(\theta) = 0$, for $a = 2, \ldots, d$, in which case θ_1 is orthogonal to the remaining parameters, then the solutions of (3.13) are

$$\pi(\theta) \propto i_{11}^{1/2}(\theta) g(\theta_2, \ldots, \theta_d), \tag{3.14}$$

where $g(\cdot)$ is an arbitrary function of its arguments (Tibshirani, 1989; Nicolaou, 1993). Although $\pi(\theta)$ is expressed as a prior for the whole vector, its property of calibration of posterior intervals applies *only* to θ_1. It is not in general possible to find a single prior that satisfies (3.13) simultaneously for all components. Unfortunately, imposing the next-order requirement, $\Pr\{\theta_1^{(1-\alpha)}(Y) \leq \theta_1 \mid \theta\} = 1 - \alpha + O(n^{-3/2})$ does not help, as the dominant term of this expansion is a function of derivatives of the log-likelihood function (Mukerjee and Ghosh, 1997).

By matching higher-order frequentist confidence limits to posterior probability limits, Fraser *et al.* (2010b) showed that the only way to have calibration to order $O(n^{-3/2})$ is to let the prior depend on the data. Data-dependent priors also featured in Box and Cox (1964) for a transformed regression model, and in Wasserman (2000) for mixture models. The use of these "priors" makes clear that the probability statements obtained from a Bayesian argument have no probabilistic meaning aside from their frequentist behaviour as confidence limits. This suggests that it may be more fruitful to directly target these intervals without using prior probabilities, which is the goal of fiducial and confidence approaches.

This discussion has assumed a (finite-dimensional) parametric model; nonparametric models raise many more difficulties. Diaconis and Freedman (1986) showed that consistency of Bayes estimates in nonparametric models can depend crucially on "arbitrary details of the prior"; see also Wasserman and Robins (2000). Mukerjee (2008) described data-dependent priors for empirical likelihoods that have probability-matching properties. Consonni *et al.* (2018) reviewed objective Bayes procedures with an emphasis on estimation and on model selection. The many-normal-means problem mentioned above is often used as a prototype for nonparametric methods. Frequentist properties of Bayesian inference for the many-normal-means problem are discussed, for example, in Castillo *et al.* (2015), Bhadra *et al.* (2016), van der Pas *et al.* (2017), and Martin and Ning (2020), although not always with an emphasis on calibration.

3.4.2 Generalized Fiducial

The fiducial argument develops a statement of uncertainty about the parameter from the structure of the model. Fisher assumed that there was a function $T(Y, \theta)$ with a known distribution. Once Y is observed, he argued that the probability that had been assigned to the unknown quantity Y could be transferred to the unknown parameter θ. This probability does not obey the usual probability calculus, which led to some confusion in the literature. Stein (1959) used the normal means example in the previous section to point out a problem with the fiducial distribution.

In spite of these reservations, Hannig and co-authors have developed generalized fiducial inference intended to be useful for complex models and high-dimensional parameters. The basis of the approach is the observation that Fisher's construction of $T(Y, \theta)$ with a known distribution is well-suited to today's method of simulating random variables. Hannig (2009) takes as a starting point a data-generating equation of the form

$$Y = G(U, \theta), \tag{3.15}$$

where U has a known distribution. If this equation can be inverted as a function of Y, then for a fixed value y, we can write

$$\theta = Q_y(U). \tag{3.16}$$

The generalized fiducial distribution for θ is defined to be $Q_y(U^*)$, where U^* is an independent copy of U. The data-generating equation is also the starting point for the discussion of fiducial decision theory by Taraldsen and Lindqvist (2024).

The function G will typically not be invertible, however, as there is no *a priori* requirement that Y and θ are structurally related, as in Fraser (1966), i.e., Y and θ take values in different spaces, usually of different dimensions. In independent sampling and continuous models, where $Y \in \mathbb{R}^n$, and $\theta \in \mathbb{R}^d$, inversion would only be possible if the particular model could be reduced, for example by sufficiency, to a d-dimensional sufficient statistic. More generally, Hannig's generalized fiducial approach introduces an auxiliary function $J(\theta; y)$ and defines the fiducial density

$$r(\theta; y) \propto f(y; \theta) J(\theta; y). \tag{3.17}$$

Here $J(\theta; y)$ functions as a kind of data-dependent prior, modifying the likelihood function $f(y; \theta)$. The construction of the fiducial density is summarized in Hannig *et al.* (2016), and (3.17) is given as a "user-friendly" formula for its construction. They note

> ... $r(\theta \mid y)$ is not a conditional density in the usual sense. Instead, we are using this notation to emphasize that the fiducial density depends on the fixed observed data.

Hannig *et al.* (2016) gave conditions under which intervals for θ computed from $r(\theta; y)$ have, in the limit, the stated coverage probabilities. (Hannig *et al.* (2016) use conditional notation as in the quote above; Murph *et al.* (2024) use the notation $r_y(\theta)$.)

But, as is the case with confidence distributions and Fisher's (1930) original fiducial distribution (Fisher, 1930), the calculus of probability does not apply, and marginal fiducial densities, generalized or not, have no probability interpretation. They may lead to confidence intervals or sets of confidence intervals that can be verified in simulations to have the stated coverage, but this needs to be checked in each problem.

A strength of the generalized fiducial approach is that by emphasizing the data-generating equation, it forces users to think about the structure of the model being fitted. It also lends itself to simulation. Hannig and co-authors have applied the method to a number of modern problems, including non-parametric estimation of a survival function (Cui and Hannig, 2019), high-dimensional regression (Lai *et al.*, 2015), linear mixed models (Cisewski and Hannig, 2012; Lidong *et al.*, 2008), and wavelet regression (Hannig and Lee, 2009). Hannig *et al.* (2016) is a good entry into this literature; see also Murph *et al.* (2024). The approach has also been used for models with discrete observations, although the inversion at (3.16) involves additional randomness (Hannig, 2009); the resulting confidence intervals are similar to continuity corrected intervals, at least in simple cases.

A weakness of the generalized fiducial approach is that additional devices seem to be needed in many complex problems, and the theoretical basis of these is not always clear, although each device may seem sensible in the context of the problem. Dawid's Example 5.17 (Dawid, 2024), is called there "non-partitionable", which in particular means that the fiducial solution must necessarily involve some *ad hoc* choice. He concludes "it could be argued ... that the model simply does not support fiducial inference".

Hannig *et al.* (2016) write:

> From a practical point of view, GFI is used in a way similar to the use of a posterior computed using a default (objective) prior, such as probability matching, reference, or flat prior. The main technical difference is that the objective prior is replaced by a data-dependent Jacobian.

As with objective Bayesian approaches, the resulting construct is not a probability statement about the parameter, but rather a statement whose empirical properties must be investigated.

3.4.3 Confidence Distributions

A general approach to the construction of confidence distributions was reviewed in Xie and Singh (2013), building on Cox (1958), Efron (1993), Schweder and Hjort (2002), and Xie *et al.* (2011), and reviewed in Thornton and Xie (2024)

Xie and Singh (2013, Definition 1) used the definition of a confidence distribution of Schweder and Hjort (2002): any function on the parameter space and sample space that is monotone increasing between 0 and 1 as a function of the parameter, and has a uniform $(0, 1)$ distribution as a function of the data. More precisely, they write

> A function $H_n(\cdot) = H_n(\mathbf{x}, \cdot)$ on $\mathcal{X} \times \Theta \to [0, 1]$ is called a confidence distribution for a parameter θ, if R1) For each given $\mathbf{x} \in \mathcal{X}$, $H_n(\cdot)$ is a cumulative distribution function on Θ; R2) At the true parameter value $\theta = \theta_0$, $H_n(\theta_0) \equiv H_n(\mathbf{x}, \theta_0)$, as a function of the sample $\mathbf{x}$, follows the uniform distribution $U[0, 1]$.

In the notation of Xie and Singh (2013), $\mathbf{x} = (x_1, \ldots, x_n)$ is a sample, and the dependence on n is introduced in order to extend the exact requirement of uniformity to an approximate one, with the approximation error (hopefully) decreasing as n increases.

This definition is agnostic about the source of the function H_n, allowing as well that there may be more than one such function. As several discussants of Xie and Singh (2013) point out, this vagueness in definition makes it difficult to distinguish a confidence distribution from a fiducial distribution, beyond the almost tautological point that confidence distributions are meant to be built up from sets of nested confidence intervals, derived by some other means. The parameter of the confidence distribution in Xie and Singh (2013) is one-dimensional, although even with this restriction confidence distributions cannot be manipulated like probability distributions. Schweder and Hjort (2013) noted that a confidence distribution for μ in a $N(\mu, 1)$ model cannot be transformed into a confidence distribution for $|\mu|$, for example.

Schweder and Hjort (2002; 2016, Chapter 6) discussed the close connection between confidence distributions and Fisher's fiducial approach. They proposed, in most cases, to seek confidence distributions for single parameters of interest, which they call focus parameters, often using as the basic quantity the profile log-likelihood ratio: in the notation of Section 3.4.1, with a model $f(y; \theta)$ and focus parameter $\psi = \psi(\theta)$, they proposed to base confidence curves for ψ on an identified pivotal quantity, i.e., a function of ψ and the data with a known distribution, and in regular statistical models they emphasize the pivotal

$$D(\psi) = 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\}, \quad \hat{\theta}_\psi = \arg\sup \ell(\theta)_{\psi(\theta)=\psi}, \tag{3.18}$$

where $\ell(\theta) = \ell(\theta; y) \propto f(y; \theta)$ is the log-likelihood function. This pivotal quantity has, at least approximately, a χ^2_1 distribution, when θ is the data-generating value, subject to some regularity conditions on the model.

To extend the range of problems that can be considered, Schweder and Hjort (2016, Chapters 3 and 4) use pivotal quantities other than (3.18) where the problem demands it. Examples include problems where the maximum likelihood estimate might be on the boundary of the parameter space, or problems with increasing numbers of nuisance parameters that can be handled by finding an appropriate marginal or conditional likelihood from which to construct a pivotal quantity. For example in the many-normal-means problem, if the target parameter is $\Sigma\theta_i^2$, then a preliminary reduction to Σy_i^2 can be used to construct a confidence distribution. While the range of problems treated is impressively large, the approach does not in general avoid an arbitrariness similar to that embodied in the definition in Xie and Singh (2013) quoted above.

When θ is a d-dimensional parameter, the χ^2_d approximation to the distribution of $D(\theta) = 2\{\ell(\hat{\theta}) - \ell(\theta)\}$ can be used to define a d-dimensional confidence distribution. Schweder and Hjort (2016, Chapter 9) note that "reducing the full confidence distribution to marginal confidence distributions for the parameters of interest ... cannot be done by integrating the confidence density", but go on to note that this might work approximately – "Whether the approximation is satisfactory must be checked, usually by simulation". This is consistent with the approach taken in generalized fiducial inference (Cisewski and Hannig, 2012; Hannig *et al.*, 2016).

A special issue on confidence distributions was published in 2018 by the *Journal of Statistical Planning and Inference* and includes many extensions of the methodology. See also Nadarajah *et al.* (2015) for an overview and several examples.

3.4.4 Significance Functions

Significance functions, or p-value functions, are essentially confidence distributions by a different name, but with an explicit focus on a single parameter. Fraser (1991) sets out the basic idea and emphasizes the link to the likelihood function. Just as the likelihood function is not a density for a parameter, so the significance function is not a distribution for the parameter, but rather a way of describing what values of the parameter are consistent with the data.

Fraser and Reid (1993), Fraser and Reid (1995) and Fraser (2017) combined the notion of significance function with a particular construction of that function. In a model for data $y \in \mathbb{R}^n$, with parameter $\theta \in \mathbb{R}^d$, and a parameter of interest $\psi \in \mathbb{R}$, as with confidence distributions and fiducial distributions, a pivotal quantity forms the basis for the significance function. In Fraser et al. (1999), the construction of the pivotal quantity first requires a dimension reduction from n to d, which is based on a local location model, generalizing an argument in Fraser (1964). This reduction is discussed in Davison and Reid (2024).

To motivate this reduction, consider Fisher's fiducial equation (3.3), but now using y in place of T:

$$\mathrm{d}f = -\frac{\partial}{\partial\theta}F(y;\theta)\mathrm{d}\theta = f(y;\theta)\frac{-\partial F(y;\theta)/\partial\theta}{f(y;\theta)} = L(\theta; y)\left|\frac{\mathrm{d}y}{\mathrm{d}\theta}\right|, \tag{3.19}$$

where $\mathrm{d}y/\mathrm{d}\theta = -F_{;\theta}(y;\theta)/F_{y;}(y;\theta)$ is the change in y with respect to θ, when the value of the distribution function is held fixed (and the abbreviated subscript notation is used for the derivative with respect to y or θ). This is generalized to the (n, d) setting as

$$\frac{\mathrm{d}y}{\mathrm{d}\theta} = \begin{pmatrix} V_1(\theta) \\ \vdots \\ V_n(\theta) \end{pmatrix} = V(\theta), \tag{3.20}$$

with each d-vector $V_i(\theta) = \mathrm{d}y_i/d\theta^{\mathrm{T}}$, obtained from the cumulative distribution function of the component y_i. (The determinant in equation (3.19) is then $|V^{\mathrm{T}}(\theta)V(\theta)|^{1/2}$.) The generalized fiducial density of Murph *et al.* (2024, Theorem 13.1) has a similar form, although they allow the "determinant-like object", here denoted $|dy/d\theta|$, to be defined in different ways.

The reduction to a d-dimensional statistical model uses the matrix $V(\hat{\theta})$ to define a tangent exponential model with canonical parameter $\varphi(\theta)$ (Reid and Fraser, 2010; Davison and Reid, 2024):

$$f_{\mathrm{TEM}}(s;\theta) = \exp\{\varphi^{\mathrm{T}}(\theta)s(y) + \ell(\theta; y^0)\}h(s). \tag{3.21}$$

In (3.21) $\ell(\theta; y^0)$ is the log-likelihood function at the observed data point y^0, and

$$\varphi(\theta) = \varphi(\theta; y^0) = \frac{\partial \ell(\theta; y^0)}{\partial V} = \sum_{i=1}^{n} \frac{\partial \ell(\theta; y^0)}{\partial y_i} V_i, \tag{3.22}$$

where the last equality assumes the components $y_1, \ldots, y_n$ are independently distributed. The model (3.21) is the basis for the development of an approximation to a significance function for ψ. The saddlepoint approximation to the density of s in an exponential family model can be applied to the density in equation (3.21), following arguments very similar to those developed in Barndorff-Nielsen (1990), and the resulting conclusion is that a modified version of the log-likelihood ratio statistic is a suitable one-dimensional pivotal quantity.

Let $r = r(\psi; y) = \pm\{2D(\psi)\}^{1/2}$, where $D(\psi)$ is defined at (3.18), and the sign on r is the same as the sign of $\hat{\psi} - \psi$. As the limiting distribution of D is χ^2_1, the limiting distribution of r is standard normal. Thus, an approximate significance function for ψ is $\Phi\{r(\psi; y^0)\}$. This approximation is accurate to $O(n^{-1/2})$, and is identical to the deviance-based approximate confidence distribution (3.18) of Schweder and Hjort (2016). A more accurate approximate significance function is $\Phi\{(r^*(\psi; y^0)\}$, where

$$r^*(\psi; y) = r + \frac{1}{r} \log \frac{Q}{r}, \tag{3.23}$$

and $Q = Q(\psi; y^0)$ is computed from $\ell(\theta; y^0)$ and $\varphi(\theta; y^0)$ in (3.21), see Davison and Reid (2024) for an explicit expression.

As a function of y with ψ fixed, the distribution of $\Phi(r^*)$ is $U(0, 1)$ under the model, to $O(n^{-3/2})$, so provides a confidence distribution in the sense of Schweder and Hjort (2002) and Xie and Singh (2013). The details of the construction of the tangent exponential model, and the normal approximation to the distribution of $r^*(\psi; y)$ under the model $f(y; \theta)$ are given in Davison and Reid (2024), building on Brazzale *et al.* (2007, Section 8.6) and a series of papers by Fraser and others.

The inferential argument that $r^*(\psi; y)$ is an appropriate pivotal quantity is more difficult and relies on showing that it emerges from conditioning on approximately ancillary statistics, first for θ in the full model, and then for ψ in the reduced model and under the constraint that ψ is fixed but θ is otherwise unconstrained. The argument is sketched in

Reid and Fraser (2010) and Fraser et al. (1999), and questions of uniqueness are considered in Fraser et al. (2010a).

3.4.5 Belief Functions and Inferential Models

A different approach to describing probability for unknown quantities is to change the formulation of probability itself. An early version of this is the development of upper and lower probabilities in Dempster (1966), extended to belief functions in the Dempster–Shafer theory (Shafer, 1976). Helpful overviews of this theory are provided in Wasserman (1990) and Dempster (2008); see also Shafer (2024) and Gong (2024). By assigning probabilities to sets of events, and defining upper and lower probabilities, the theory permits a conclusion of uncertainty for the truth of an event. This is related to the theory of imprecise probabilities developed in Walley (1990). While this theory is philosophically appealing from some points of view, it has not had broad impact on statistical practice.

One modern version of this theory is the development of inferential models (Martin and Liu, 2016). These models are developed in a manner reminiscent of Hannig's generalized fiducial distribution (Hannig, 2009). The latter is computed using a data-generating equation $Y = G(U, \theta)$ and an independently generated copy of U^*, as in Section 3.4.2. The inferential model approach proposes using a random set to predict U; this random set is converted to a belief function. Dawid (2024, Section 5.9) describes how Hannig's generalized fiducial inference and Martin and Liu's inferential models build on the theory of belief functions.

The derivation of the inferential model is guided by the requirement that the resulting inference is both valid and efficient, properties which are analogous to coverage and length of a confidence interval; definitions are given in Martin and Liu (2016, Chapter 4). Martin (2019) argued that it is not possible to obtain valid statistical inference in general without using a theory of non-additive probability, such as that provided by inferential models. Balch *et al.* (2019) argued that Bayesian and fiducial methods of inference are necessarily at risk of being invalid, because they are based on the usual probability calculus, rather than a system of belief functions.

Liu and Martin (2024) link their development of inferential models to the theory of possibility measures, which involve a particular type of belief function.

3.5 Comparisons

As developments in statistical science turned increasingly to implementation of methods, the historical tension among Bayesian, frequentist and fiducial approaches was not resolved so much as ignored. The current study of the theory described above is ongoing, and the test of time may be which versions are most flexible and computationally feasible in complex models.

Some points of comparison of the approaches include the nature of the probability statements sought, the adjustment or allowance for nuisance parameters, and the potential for calibration.

Even the nature of the probability statement intended is not always clear from what is written. There seems to be a presumption in the application of Bayesian methods in practice that the posterior probability has an epistemic interpretation, but the marginal posterior probability for a quantity of interest has no such interpretation unless the prior does, and various proposed noninformative or objective priors have no such meaning. Berger (2006) provided a series of philosophical viewpoints around Bayesian analysis using objective priors which emphasizes their practical importance over their interpretation. The weakest of these is his fourth: "Objective Bayesian analysis is simply a collection of *ad hoc* but useful methodologies for learning from data"; this is close to the viewpoint outlined by Fraser (2011). Berger *et al.* (2010) discussed the formal probability interpretation of a posterior distribution obtained using the reference prior of Bernardo (1979).

Empirical Bayes inference (Efron, 2010; 2024) is a seemingly firmly frequentist approach to inference, estimating from the sample either the marginal density $m(y)$ or the prior density $\pi(\theta)$: Efron's f-estimates and g-estimates, respectively. Much of the literature on empirical Bayes methodology emphasizes its use in nonparametric settings and establishing concentration rates for posteriors (see, e.g. Donnet *et al.* (2018); Petrone *et al.* (2014); Zhang and Gao (2020)). Efron (2024, Sections 2.5, 2.6) treats the estimated posterior as a distribution for θ; see, e.g., his equation (2.95), but it is not clear to me how much is assumed in his phrase "Taking $\hat{g}$ literally".

Generalized fiducial distributions are described as a pragmatic way to generate (sets of) confidence intervals, for example Hannig *et al.* (2016, p. 1348) write "GFD ... should be viewed as a distribution estimator ... of the fixed true parameter. To validate this distribution estimator in a specific example, we then typically demonstrate good small sample performance by simulation and prove good large sample properties by asymptotic theorems". This suggests the operational definition of a method of inference that is assessed by its calibration properties, as described in Reid and Cox (2015). Significance functions are used exactly in the same way in Fraser (1991), as a summary of the information about a parameter available from the data, with properties of the summary to be assessed by calibration.

Schweder and Hjort (2016, p. 4) described a confidence distribution as a "gold standard for epistemic probability distributions in science". They emphasized that confidence densities cannot be treated by the usual probability calculus, but they ascribed this to the inapplicability of the usual calculus to epistemic probability. In their construction of confidence distributions, they ensured that its properties in repeated sampling, i.e. its aleatory or empirical properties, are guaranteed. The epistemic nature of the confidence distribution is then simply a statement about the uncertainty of any conclusions we may draw from the data at hand. It is not clear what is gained by describing the confidence distribution as an epistemic probability, which seems to put it on the same foundational footing as a posterior probability. But their approach does satisfy the requirement described in Reid and Cox (2015) that any inferential procedure be treated as a measuring device, and its properties assessed empirically.

The approach to inference based on confidence distributions outlined by Xie and Singh (2013) is similarly proposed as a method for generating reasonable inferences, "a redefinition of the confidence distribution as a *purely frequentist concept*", although in the same section they describe a confidence distribution as a "probability distribution on the parameter space", while making clear in later sections that this distribution does not obey the usual rules of probability calculus. More recent work emphasizes techniques for combining confidence distributions; see, e.g., Shen *et al.* (2020).

Thornton and Xie (2024) argue that the theory of confidence distributions can serve as a bridge between Bayes, frequentist and fiducial approaches.

> Regardless of whether the setting is parametric or nonparametric, normal or non-normal, exact or asymptotic, CD inference is possible as long as one can create confidence intervals (or regions) of all levels for the parameter of interest.

This is essentially the description of a confidence distribution as a summary statistic, or function, and is agnostic about the method used to derive it. From that point of view, the difficulties inherent in eliminating nuisance parameters, for example, have been solved by the method used to create the confidence distribution, but there remains no general theory for ensuring that this is possible, and the interpretation of the result is firmly aleatory.

The conclusions from inferential models are intended to be epistemic: "The inferential output should be interpreted as a belief probability, or a range of belief probabilities. That is, there is no notion of frequencies in its interpretation, it is simply a measure of evidence in data" (Martin and Liu, 2016, p. 13). In this sense, inferential models and belief functions provide a different axiomatic theory, seemingly contradicting the statement of Schweder and Hjort above. However, Martin and Liu (2016) emphasized what they call validity as a necessary component of inferential models, which seems to me to put priority on the calibration of the resulting inference.

Indeed all the methods outlined in Section 3.4 are generally described by nearly all their proponents as needing to be assessed by their frequentist performance. The only method for which such assessment is seen as irrelevant is a completely subjective Bayes approach, although even there the possibility may be allowed that for any particular application of the method, the inference obtained is not very much influenced by the choice of prior distribution (Goldstein, 2006).

There is also general agreement among the proponents of approaches of Section 3.4 that inference for a parameter of interest cannot be obtained by marginalizing over a joint posterior, or confidence, or fiducial, distribution. The theory of reference priors, for example, is very clear about this (Berger and Bernardo, 1992). Schweder and Hjort (2016, Chapter 2.4) emphasized the construction of confidence distributions for a parameter of interest, which they call a focus parameter, by reduction of the data in advance of the construction. This typically proceeds by finding an exact or approximate pivotal quantity.

In more general settings, but with smooth models, approximate pivotal quantities can be obtained by using the classical summary statistics from likelihood inference. Xie and Singh (2013) often used the Wald statistic $(\hat{\psi} - \psi)/\hat{\sigma}_{\psi}$, which under repeated sampling from the model follows an approximate standard normal distribution, subject to regularity conditions. Schweder and Hjort (2016) usually use the asymptotically equivalent pivotal $D(\psi) = 2\{\ell(\hat{\psi}, \hat{\lambda}) - \ell(\psi, \hat{\lambda}_{\psi})\}$; its approximate sampling distribution is χ^2_1. In non-regular models special arguments are needed to find a suitable pivotal quantity. Hannig's approach to generalized fiducial distributions sometimes reduces the problem to a pivotal quantity first, and sometimes marginalizes a multi-parameter generalized fiducial distribution, but then checks that the resulting inference is approximately valid.

As described in Davison and Reid (2024), the theory of significance functions identifies a particular pivotal quantity, $r^*(\psi; y)$, based on arguments derived from higher-order asymptotic expansions of likelihood-based quantities. An advantage of insisting on this, beyond the possibly improved distributional approximation, is that approximate elimination of nuisance parameters is built into its construction. In effect, nuisance parameters are integrated out of the model using Laplace approximation, and the validity of this integration step also rests on higher-order approximations to likelihood-based quantities. In this way, the significance function approach removes the *ad hoc* choices involved in developing confidence distributions or generalized fiducial distributions. Fraser was convinced that the pivotal quantity $r^*(\psi; y)$ was essentially unique, but the arguments in support of this still seem to me to be somewhat elusive.

Higher-order asymptotic theory assumes a high degree of regularity in the models, whereas with non-regular problems there is often a clever construction that can be exploited. For example, Cunen *et al.* (2018) proposed two methods for constructing confidence distributions for change-points, one based on a test for no change, and another based on the simulated distribution of the profile log-likelihood ratio statistic. Cui and Hannig (2019) showed how to invert a pair of data-generating equations for censored survival data to obtain a non-parametric generalized fiducial distribution for the survival function. I am not sure if either of these problems can be readily handled using the r^* construction.

The theory of inferential models has both a method for conditioning on auxiliary variables, to reduce the dimension of the data to that of the parameter, and a method for marginalizing over auxiliary variables to eliminate nuisance parameters; see Martin and Liu (2016, Chapters 6 and 7).

3.6 Conclusion

This is necessarily a very selective overview, and the quotes from other papers in this volume and in the literature can give only a limited view. The emphasis on distributions for parameters means that nonparametric models are not addressed. Vovk (2024) describes nonparametric fiducial methods for predicting future observations, and shows how a generalization of this is connected to the theory of conformal prediction. Taraldsen and Lindqvist (2024) discuss fiducial decision theory; the emphasis is on an action, or decision, made on the basis of some data, model and loss function.

Sen *et al.* (2024) study an enormously complex parametric model, which could be viewed as practically non-parametric, arguing that a Bayesian approach to inference is more likely to be reproducible. Implicitly, the high-dimensional posterior distribution is integrated to obtain marginal posterior distributions; it seems likely that this is subject to the Stein phenomenon described in Section 3.4.1. Similarly, the massive data sets envisaged in Hector *et al.* (2024) require more emphasis on computation than interpretation; the goal in that paper is to construct efficient estimators of parameters that borrow strength from datasets of similar structure, which requires first assessing which datasets these might be. The focus on individual parameters and their potential distribution is of less interest in these complex models.

Berger (2022) categorized several types of frequentists, or several flavours of frequentism, and Berger (2006) described four versions of objective Bayes. This introduction, and volume, confirms that there are many approaches to find distributions for parameters, and perhaps also confirms that the collective noun for statisticians is *a disagreement.*

It is perhaps natural, given the subtlety of the concepts involved, that there continues to be disagreement, or at least not convergence, among various approaches to the interpretation of distributions for parameters. The true test of statistical theory, though, is its utility and reliability for applications, and often the disagreement isn't as stark, at least in problems with relatively well-understood models and modest amounts of data.

A recent series of papers on the estimation of the probability of collisions among orbiting satellites illuminates some of these points. The calculations are essentially those for inference on the length of a multivariate normal vector, (Stein, 1959), discussed in Section 3.4.1 and in Dawid (2024, Section 5.7.2). Balch *et al.* (2019) note the mis-calibration of the Bayes posterior, and state a general principle which they call the false confidence theorem. Cunen *et al.* (2021) pointed out that the problem is solved by first reducing to the obvious pivotal quantity, the length of the response vector, before constructing a confidence distribution. Martin *et al.* (2021) responded that their theorem continues to be valid, as confidence distributions cannot be manipulated as if they were probability distributions. As far as practical solution of the problem of satellite collisions, though, the approach of Cunen *et al.* (2021) or the improvement developed in Elkantassi and Davison (2022) based on Section 3.4.4 seem quite satisfactory.

We rely on foundations and theory to provide a basis for developing solutions to new practical problems; without this each application would require some *ad hoc* solution anew. The papers in this volume illustrate the subtleties in foundational concepts and are at the same time motivated by the desire to improve statistical practice.

Bibliography

Balch, M. S., Martin, R. and Ferson, S., 2019. Satellite conjunction analysis and the false confidence theorem. *Proceedings of the Royal Society A*, *475*, p. 120180565. https://doi.org/10.1098/rspa.2018.0565.

Barndorff-Nielsen, O. E., 1990. Approximate interval probabilities. *Journal of the Royal Statistical Society: Series B*, *52*, pp. 485–496.

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, *53* and *54*, pp. 370–418; 296–325. Reprinted in *Biometrika*, *45*, 1958, pp. 293–315.

Berger, J. O., 2006. The case for objective Bayes analysis. *Bayesian Analysis*, *1*, pp. 385–402.

Berger, J. O., 2022. Four types of frequentism and their interplay with Bayesianism (with discussion). *New England Journal of Statistics in Data Science*, *1*, pp. 126–137.

Berger, J. O. and Bernardo, J. M., 1992. On the development of reference priors (with discussion). In *Bayesian Statistics 4*, eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp. 35–60. Oxford: Oxford University Press.

Berger, J. O., Bernardo, J. M. and Sun, D., 2010. The formal definition of reference priors. *Annals of Statistics*, *37*, pp. 905–938.

Berger, J. O., Garcia-Donato, G., Moreno, E. and Pericchi, L., 2024. Objective Bayesian testing and model uncertainty. In *A Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 132–160. Boca Raton: Chapman and Hall/CRC.

Berger, J. O., Insua, D. R. and Ruggeri, F., 2000. Bayesian robustness. In *Robust Bayesian Analysis*, eds D. R. Insua and F. Ruggeri, Lecture Notes in Statistics, vol. 152, pp. 1–32. New York: Springer.

Bernardo, J. M., 1979. Reference posterior distributions for Bayesian inference (with discussion). *Journal of the Royal Statistical Society: Series B*, *41*, pp. 113–147.

Bhadra, A., Datta, J., Polson, N. G. and Willard, B., 2016. Default Bayesian analysis with global-local shrinkage priors. *Biometrika*, *103*, pp. 955–969.

Box, G. E. P. and Cox, D. R., 1964. An analysis of transformations (with discussion). *Journal of the Royal Statistical Society: Series B*, *26*, pp. 211–246.

Brazzale, A. R., Davison, A. C. and Reid, N., 2007. *Applied Asymptotics.* Cambridge: Cambridge University Press.

Castillo, I., Schmidt-Hieber, J. and van der Vaart, A., 2015. Bayesian linear regression with sparse priors. *Annals of Statistics*, *43*, pp. 1986–2018.

Cisewski, J. and Hannig, J., 2012. Generalized fiducial inference for normal linear mixed models. *Annals of Statistics*, *40*, pp. 2101–2127.

Consonni, G., Fouskakis, D., Liseo, B. and Ntzoufras, I., 2018. Prior distributions for objective Bayesian analysis. *Bayesian Analysis*, *13*, pp. 627–679.

Cox, D. R., 1958. Some problems connected with statistical inference. *Annals of Mathematical Statistics*, *29*, pp. 375–372.

Cox, D. R., 2016. Some pioneers of modern statistical theory: A personal reflection. *Biometrika*, *103*, pp. 747–759.

Cui, Y. and Hannig, J., 2019. Nonparametric generalized fiducial inference for survival functions under censoring. *Biometrika*, *106*, pp. 501–518.

Cunen, C., Hermansen, G. and Hjort, N. L., 2018. Confidence distributions for change-points and regime shifts. *Journal of Statistical Planning and Inference*, *195*, pp. 14–34.

Cunen, C., Hjort, N. L. and Schweder, T., 2021. Confidence in confidence distributions! *Proceedings of the Royal Society A*, *476*, p. 20190781.

Datta, G. S. and Mukerjee, R., 2004. *Probability Matching Priors: Higher Order Asymptotics*. New York: Springer.

Davison, A. C. and Reid, N., 2024. The tangent exponential model. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 210–237. Boca Raton: Chapman and Hall/CRC.

Dawid, A. P., 2024. Fiducial inference: Then and now. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 83–105. Boca Raton: Chapman and Hall/CRC.

Dawid, A. P., Stone, M. and Zidek, J. V., 1973. Marginalization paradoxes in Bayesian and structural inference. *Journal of the Royal Statistical Society: Series B*, *35*, pp. 189–233.

Dempster, A. P., 1966. New methods for reasoning towards posterior distributions based on sample data. *Annals of Mathematical Statistics*, *37*, pp. 355–374.

Dempster, A. P., 2008. The Dempster–Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, *48*, pp. 365–377.

Diaconis, P. and Freedman, D., 1986. On the consistency of Bayes estimates. *Annals of Statistics*, *14*, pp. 1–26.

Donnet, S., Rivoirard, V., Rousseau, J. and Scricciolo, C., 2018. Posterior concentration rates for empirical Bayes procedures with applications to Dirichlet process mixtures. *Bernoulli*, *24*, pp. 231–256.

Efron, B., 1993. Bayes and likelihood calculations from confidence intervals. *Biometrika*, *80*, pp. 3–26.

Efron, B., 2010. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge: Cambridge University Press.

Efron, B., 2024. Empirical Bayes: Concepts and methods. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 8–34. Boca Raton: Chapman and Hall/CRC.

Elkantassi, S. and Davison, A. C., 2022. Space oddity? A statistical formulation of conjunction assessment. *Journal of Guidance, Control, and Dynamics*, *45*, pp. 2258–2274.

Evans, M. and Moshonov, H., 2006. Checking for prior-data conflict. *Bayesian Analysis*, *1*, pp. 893–914.

Fisher, R. A., 1930. Inverse probability. *Proceedings of the Cambridge Philosophical Society*, *xxvi*, pp. 528–535.

Fisher, R. A., 1956. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver & Boyd.

Fraser, A. M., Fraser, D. A. S. and Staicu, A.-M., 2010a. Second order ancillary: A differential view with continuity. *Bernoulli*, *16*, pp. 1208–1223.

Fraser, D. A. S., 1964. Local conditional sufficiency. *Journal of the Royal Statistical Society: Series B*, *26*, pp. 52–62.

Fraser, D. A. S., 1966. Structural probability and a generalization. *Biometrika*, *53*, pp. 1–9.

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

Fraser, D. A. S., 1991. Statistical inference: Likelihood to significance. *Journal of the American Statistical Association*, *86*, pp. 258–265.

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

Fraser, D. A. S., 2017. p-values: The insight to modern statistical inference. *Annual Review of Statistics and Its Application*, *4*, pp. 1–14.

Fraser, D. A. S. and Reid, N., 1992. Strong matching of frequentist and Bayesian parametric inference. *Journal of Statistical Planning and Inference*, *103*, pp. 263–285.

Fraser, D. A. S. and Reid, N., 1993. Third order asymptotic models: Likelihood functions leading to accurate approximation to distribution functions. *Statistica Sinica*, *3*, pp. 67–82.

Fraser, D. A. S. and Reid, N., 1995. Ancillaries and third order significance. *Utilitas Mathematica*, *47*, pp. 33–53.

Fraser, D. A. S., Reid, N., Marras, E. and Yi, G. Y., 2010b. Default priors for Bayesian and frequentist inference. *Journal of the Royal Statistical Society: Series B*, *72*, pp. 631–654.

Fraser, D. A. S., Reid, N. and Wu, J., 1999. A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika*, *86*, pp. 249–264.

Gelman, A., Jakulin, A., Pittau, M. G. and Su, Y. S., 2008. A weakly informative default prior distribution for logistic regression models. *Annals of Applied Statistics*, *2*, pp. 1360–1383.

Goldstein, M., 2006. Subjective Bayesian analysis: Principles and practice. *Bayesian Analysis*, *1*, pp. 403–420.

Gong, R., 2024. Dempster-Shafer theory for statistical inference. In *Handbook of Bayesian Fiducial and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie. Boca Raton: Chapman and Hall/CRC. pp. 300–322.

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

Hannig, J., Iyer, H., Lai, R. C. S. and Lee, T. C. M., 2016. Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, *111*, pp. 1346–1361.

Hannig, J. and Lee, T. C. M., 2009. Generalized fiducial inference for wavelet regression. *Biometrika*, *96*, pp. 847–860.

Hector, E.C., Tang, L., Zhou, L. and Song, P. X.-K., 2024. Data integration and model fusion in the Bayesian, fiducial and frequentist framerwork. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 238–263. Boca Raton: Chapman and Hall/CRC.

Heinrich, J., 2006. The Bayesian approach to setting limits – what to avoid. In *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, eds L. Lyons and M. K. Ünel, pp. 98–101. London: Imperial College Press.

Jeffreys, H., 1961. *Theory of Probability* (3rd ed.). Oxford: Oxford University Press.

Kass, R. E. and Wasserman, L., 1996. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, *91*, pp. 1343–1370.

Kleijn, B. J. K. and van der Vaart, A. W., 2012. The Bernstein-Von-Mises theorem under model misspecification. *Electronic Journal of Statistics*, *6*, pp. 354–381.

Kline, D., Li, Z., Chu, Y., Wakefield, J., Miller, W. C., Turner, A. N. and Clark, S. J., 2021. Estimating seroprevalence of SARS-CoV-2 in Ohio: A Bayesian multilevel poststratification approach with multiple diagnostic tests. *Proceedings of the National Academy of Sciences*, *118*. https://doi.org/10.1073/pnas.2023947118.

Lai, R. C. S., Hannig, J. and Lee, T. C. M., 2015. Generalized fiducial inference in ultrahigh-dimensional regression. *Journal of the American Statistical Association*, *110*, pp. 760–772.

Lang, J. B., 2024. Slicing and dicing a path through the fiducial forest. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 323–343. Boca Raton: Chapman and Hall/CRC.

Lehmann, E. L., 1993. The Fisher, Neyman-Pearson theories of testing hypotheses: One test or two? *Journal of the American Statistical Association*, *88*, pp. 1242–1249.

Lemoine, N., 2019. Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analysis. *Oikos*, *128*, pp. 912–928.

Lidong, E., Hannig, J. and Iyer, H. K., 2008. Fiducial intervals for variance components in an unbalanced two-component normal mixed linear model. *Journal of the American Statistical Association*, *103*, pp. 854–864.

Liu, C. and Martin, R., 2024. Inferential models and possibility measures. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 344–363. Boca Raton: Chapman and Hall/CRC.

Martin, R., 2019. False confidence, non-additive beliefs, and valid statistical inference. *International Journal of Approximate Reasoning*, *113*, pp. 39–73.

Martin, R., Balch, M. S. and Ferson, S., 2021. Response to the comment confidence in confidence distributions! *Proceedings of the Royal Society A*, *477*, p. 20200579.

Martin, R. and Liu, C., 2016. *Inferential Models: Reasoning with Uncertainty*. Boca Raton: Chapman & Hall/CRC.

Martin, R. and Ning, B., 2020. Empirical priors and coverage of posterior credible sets in a sparse normal mean model. *Sankhya A*, *82*, pp. 477–498.

Meng, X.-L., 2024. A BFFer's exploration with nuisance constructs: Bayesian p-value, H-likelihood, and Cauchyanity. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 161–187. Boca Raton: Chapman and Hall/CRC.

Mukerjee, R., 2008. Data-dependent probability matching priors for empirical and related likelihoods. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, eds B. Clarke and S. Ghosal, pp. 60–70. Hayward: IMS Collections.

Mukerjee, R. and Ghosh, M., 1997. Second-order probability matching priors. *Biometrika*, *84*, pp. 970–975.

Murph, A., Hannig, J. and Williams, J. P., 2024. Introduction to generalized fiducial inference. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 276–299. Boca Raton: Chapman and Hall/CRC.

Nadarajah, S., Bityukov, S. and Krasnikov, N., 2015. Confidence distributions: A review. *Statistical Methodology*, *22*, pp. 23–46.

Neyman, J., 1937. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society A*, *236*, pp. 333–380.

Nicolaou, A., 1993. Bayesian intervals with good frequentist properties in the presence of nuisance parameters. *Journal of the Royal Statistical Society: Series B*, *55*, pp. 377–390.

van der Pas, S., Szabo, B. and van der Vaart, A., 2017. Uncertainty quantification for the horseshoe (with discussion). *Bayesian Analysis*, *12*, pp. 1221–1274.

Peers, H. W., 1965. On confidence points and Bayesian probability points in the case of several parameters. *Journal of the Royal Statistical Society: Series B*, *27*, pp. 9–16.

Petrone, S., Rousseau, J. and Scricciolo, C., 2014. Bayes and empirical Bayes: Do they merge? *Biometrika*, *101*, pp. 285–302.

Reid, N. and Cox, D. R., 2015. On some principles of statistical inference. *International Statistical Review*, *83*, pp. 293–308.

Reid, N. and Fraser, D. A. S., 2010. Mean loglikelihood and higher-order approximations. *Biometrika*, *97*, pp. 159–170.

Reid, N. and Sun, Y., 2010. Assessing sensitivity to priors using higher order approximations. *Communications in Statistics—Theory and Methods*, *39*, pp. 1373–1386.

Savage, L. J., 1954. *The Foundations of Statistics*. New York: Wiley.

Schervish, M. J., 1995. *Theory of Statistics*. New York: Springer.

Schweder, T. and Hjort, N. L., 2002. Confidence and likelihood. *Scandinavian Journal of Statistics*, *29*, pp. 309–332.

Schweder, T. and Hjort, N. L., 2013. Discussion of "Confidence distribution, the frequentist distribution estimator of a parameter". *International Statistical Review*, *81*, pp. 56–68.

Schweder, T. and Hjort, N. L., 2016. *Confidence, Likelihood, Probability: Inference with Confidence Distributions*. Cambridge: Cambridge University Press.

Sen, D., Papamarkou, T. and Dunson, D., 2024. Bayesian neural networks and dimensionality reduction. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 188–209. Boca Raton: Chapman and Hall/CRC.

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

Shafer, G., 2024. How the game-theoretic foundation for probability resolves the Bayesian vs. frequentist standoff. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 264–275. Boca Raton: Chapman and Hall/CRC.

Shen, J., Liu, R. Y. and Xie, M., 2020. iFusion: Individualized fusion learning. *Journal of the American Statistical Association*, 115, pp. 1251–1267. https://doi.org/10.1080/01621459.2019.1672557.

Stein, C., 1959. An example of wide discrepancy between fiducial and confidence intervals. *Annals of Mathematical Statistics*, *30*, pp. 877–880.

Stigler, S., 1986. Laplace's 1774 memoir on inverse probability. *Statistical Science*, *1*, pp. 359–363.

Stigler, S., 2013. The true title of Bayes's essay. *Statistical Science*, *28*, pp. 283–288.

Taraldsen, G. and Lindqvist, B. H., 2024. Fiducial inference and decision theory. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 381–395. Boca Raton: Chapman and Hall/CRC.

Thornton, S. and Xie, M., 2024. Bridging Bayesian, frequentist and fiducial inferences using confidence distributions. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 106–131. Boca Raton: Chapman and Hall/CRC.

Tibshirani, R., 1989. Noninformative priors for one parameter of many. *Biometrika*, *76*, pp. 604–608.

Vovk, V., 2024. Conformal predictive distributions: An approach to nonparametric fiducial prediction. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, eds J. Berger, X-L. Meng, N. Reid and M. Xie, pp. 364–380. Boca Raton: Chapman and Hall/CRC.

Walley, P., 1990. *Statistical Reasoning with Imprecise Probabilities*. London: Chapman & Hall.

Wasserman, L. A., 1990. Belief functions and statistical inference. *Canadian Journal of Statistics*, *18*, pp. 183–196.

Wasserman, L. A., 2000. Asymptotic inference for mixture models using data-dependent priors. *Journal of the Royal Statistical Society: Series B*, *62*, pp. 159–180.

Wasserman, L. A. and Robins, J., 2000. Conditioning, likelihood and coherence: A review of some foundational concepts. *Journal of the American Statistical Association*, *95*, pp. 1340–1346.

Xie, M. and Singh, K., 2013. Confidence distributions: The frequentist distribution estimator of a parameter. *International Statistical Review*, *81*, pp. 3–39.

Xie, M., Singh, K. and Strawderman, W. E., 2011. Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association*, *106*, pp. 320–333.

Zabell, S. L., 1992. R. A. Fisher and the fiducial argument. *Statistical Science*, *7*, pp. 369–387.

Zhang, F. and Gao, C., 2020. Convergence rates of empirical Bayes posterior distributions: A variational perspective. *arXiv e-prints*, arXiv:2009.03969.