Philosophy of Science (2024), **00**, 1–10
doi:10.1017/psa.2023.174

PSA PHILOSOPHY OF SCIENCE ASSOCIATION
PHILOSOPHY OF SCIENCE

SYMPOSIA PAPER

An Exploration of Parameter Duality in Statistical Inference

Suzanne Thornton[1,2] and Minge Xie[3]

[1]Mathematics and Statistics Department, Swarthmore College, Swarthmore, PA, USA, [2]Mathematics and Statistics Department, George Washington University, Washington, DC 20052, USA and [3]Rutgers University, New Brunswick, NJ, 08901, USA
Corresponding author: Suzanne Thornton; Email: suzannet@gwu.edu

(Received 19 April 2023; revised 04 December 2023; accepted 04 December 2023)

Abstract

Well-known debates among statistical inferential paradigms emerge from conflicting views on the notion of probability. One dominant view understands probability as a representation of sampling variability; another prominent view understands probability as a measure of belief. The former generally describes model parameters as fixed values, in contrast to the latter. We propose that there are actually two versions of a parameter within both paradigms: a fixed unknown value that generated the data and a random version to describe the uncertainty in estimating the unknown value. An inferential approach based on CDs deciphers seemingly conflicting perspectives on parameters and probabilities.

1 Introduction

Frequentist inference is connected to the logic of probability through the notion of empirical randomness. Sample estimates are useful insofar as one has a sense of the extent to which the estimator may vary from one sample to another. The bounds of a confidence interval are particular observations of random variables, where the randomness is inherited by the random sampling of data. For example, 95% confidence intervals for parameter θ can be calculated for any random sample from a normal $N(\theta, 1)$ distribution. With repeated sampling (assuming a correct model), approximately 95% of these intervals are guaranteed to yield an interval covering the fixed value of θ.

Bayesian inference produces a probability distribution for the different values of a particular parameter, even though the quality of this distribution is often difficult to assess without invoking an appeal to the notion of repeated performance. For data observed from an $N(\theta, 1)$ distribution, generating a credible interval for θ requires an assumption about the plausibility of different possible values of θ; that is, one must assume a prior (in addition to assuming the model is correct). However, depending on the context—is θ the efficacy of a newly created drug, or is θ the efficacy of a new

version of an older drug?—there may or may not be an informed choice for the prior. Without appealing to the long-run performance of the interval, how can one compare a 95% credible interval $[a, b]$ versus another 95% interval $[a', b']$ based on the same data but a different prior?

We explore a paradigm that elucidates the philosophical tie between statistical estimation and inference by targeting distribution estimators with desirable properties, that is, *confidence distribution (CD) estimators*. Although we believe this perspective has broad potential to address many challenging problems in modern statistics (e.g., Xie and Singh 2013; Reid and Cox 2015), for clarity, we restrict our attention to a simple inferential setting assuming a relationship between observed data, $x_{\text{obs}} = \{x_1, \ldots, x_n\}$, and an unknown, one-dimensional parameter, θ. Most often, this relationship is characterized through the likelihood function, but we prefer a general description of this relationship as a *data-generating model* dependent on θ; symbolically, $x_{\text{obs}} \sim \text{Model}(\theta) \stackrel{\text{def}}{=} M_\theta$. Critically, a data-generating model need not correspond to a tractable likelihood. Because model-selection criteria are not the focus of this article, we assume throughout that M_θ is correct.

The purpose of this article is not to critique any statistical inferential paradigm; however, to clarify our perspective, we present some key questions. By broadening the perspective of statistical estimation to include distribution estimators with desirable performance standards, we find possible resolutions. Ultimately, we challenge a common assertion that Bayesian and frequentist methods differ primarily in their interpretation of a model parameter and conventional adherence to a singular role for probability in statistical inference.

1.1 Statistical inference

The main distinction between Bayesian and frequentist inference is that the former requires an initial distribution estimator, a *prior*, which is refined by conditioning upon the observed data. Broadly speaking, there are two Bayesian techniques for selecting a prior. A subjective Bayesian strictly employs the prior as a representation of the degrees of an agent's belief. Subjective methods can be prohibitively challenging in practice. Objective Bayesians instead use the prior to represent the population of possible parameter values. Objective Bayesian priors need not follow probability laws and may even depend on the same data used in the data-generating model (likelihood). Frequentists do not require a prior, but as demonstrated later, they can produce data-dependent distribution functions on the parameter space.

A strong example of agreement between Bayesian and frequentist paradigms is easily identified within an objective Bayesian context. Objective Bayesianism provides statistical answers conditioned upon x_{obs} while also respecting the (frequentist) notion that statistical methodology should ensure successful conclusions most of the time. This agreement on a necessary standard of performance for a statistical method is an appeal to the notion of *calibration* (Reid and Cox 2015). Of the paradigms discussed, only subjective Bayesians may disregard calibration.

In any case, the quantification of uncertainty serves an essential purpose (although this may not be the end goal), and there emerges a dual nature in the treatment of the model parameter; that is, there may be both a random version of a parameter and a deterministic version. Through the notion of CD estimators (defined later), we find

https://doi.org/10.1017/psa.2023.174 Published online by Cambridge University Press

that certainty about a parameter θ can be expressed probabilistically even without an assumed prior distribution. This begs the question, Are epistemic meanings justifiable from calibrated methods that do not assume a (proper) prior?

2 Dualities in statistical inference

2.1 The dual role of parameters

New learners are often taught that Bayesian and frequentist methods differ in the interpretation of model parameters—a parameter is a fixed unknown quantity for frequentists but a random variable for Bayesians. This is a subtle misunderstanding that we attempt to clarify. Parameter θ represents a deterministic value, presumably having generated x_{obs}; however, it may also be associated with a random variable describing the uncertainty in determining what the particular value is. This dual role is apparent in the original manuscript of Thomas Bayes's famous theorem and is also evident in modern statistical methods reliant on computer-simulated data.

In lemma 1 of section 2, Bayes (1764) describes the following setting. Consider an arbitrary flat surface (say, a billiard table) of length 1. A first player rolls ball W along the billiard table and secretly marks the location where it stops. A second player rolls ball O n different times on the same billiard table. If the only information (i.e., data) the second player observes is a report from the first player on whether or not ball O lands to the left of ball W on each of the n trials, then what inference can the second player make concerning the location of ball W?

In modern statistical terms, Bayes's billiard table experiment assumes a $U(0,1)$ prior distribution, and the sample of observed data is realized from a binomial model. In this context, θ_0 (where ball W landed) is a realization from the prior distribution, $\theta \sim U(0,1)$, so the sample is generated from a binomial distribution with n trials and probability of success θ_0.

Bayes's scenario clearly displays two versions of the parameter at play, which Thornton and Xie (forthcoming) call the *duality of a parameter*. In application, estimating the fixed version of the parameter that generated the observed data is the main objective. We denote this target value with the symbol θ_0. There is, however, some uncertainty in exactly how to assess θ_0, given x_{obs}. To address this uncertainty, a Bayesian approach elects to work with a random representation of θ (the prior), which is then updated to produce a data-informed posterior distribution. In this approach, "[p]arameters must have a distribution describing the available information about their values ... [but] this is not a description of their variability (since they are fixed unknown quantities), [rather it is] a description of the uncertainty about their true values" (Berger et al. 2015). One may surmise that probabilistic statements about an unknown parameter (e.g., the prior or posterior) are not statements asserting the parameter is actually a random variable itself; rather, these statements describe the (un)certainty in estimating an unobserved, fixed, data-generating value.

Many modern inference problems rely on computer simulations to mimic whatever type of randomness (i.e., descriptions of [un]certainty or sample variability) is used to answer the posed inferential question. We refer to the intentional use of computer simulations to generate possible values of a random variable as *artificial randomness*. Artificial randomness is not exactly empirical, nor is it clearly

https://doi.org/10.1017/psa.2023.174 Published online by Cambridge University Press

epistemological. Indeed, the interpretation of artificial randomness is an interesting open philosophical question.

Several established statistical methods utilize artificial randomness, including approximate Bayesian computing and frequentist bootstrap methods; see, for example, Beaumont (2019) and Efron (2000). Once again, two versions of θ emerge in these methods: those values simulated via artificial randomness, say, θ^*, and a deterministic value that is connected to the observed data, θ_0. Typically, such methods are theoretically supported by a probabilistic matching condition, which can be represented as an approximate (or exact) matching between two random entities:

$$[(\theta^* - f(X))|X = x_{\text{obs}}] \approx (f(X) - \theta_0), \tag{1}$$

where θ^* is produced through (post-data) artificial randomness, and $f(X)$ is a function of the random sample of data, X, whose distribution depends on the unknown model parameter. This is a statement about the probability law governing the random sample, X (right-hand side), matching the probability law inducing the artificially random parameter values, θ^* (left-hand side).

Thornton and Xie (forthcoming) describe this matching condition more generally, across Bayesian, frequentist, and fiducial paradigms. The common element among any statistical method employing this theorem is that there are two versions of the model parameter at play. The first version is a fixed target value related to the observed sample of data. The second version is a result of artificial randomness and describes the unknown parameter probabilistically (after conditioning on x_{obs}). In this way, these complex modern inference problems mirror Bayes's original proposal, wherein the objective is a statement regarding a deterministic target, but probability is used to describe the uncertainty associated with that target, given an observed data set.

Traditionally, both Bayesians and frequentists rely on the assumption that the model is correct in their use of the likelihood function. In modern applications where no likelihood may be available, these methods differ in the extent to which they require assumptions about the underlying data-generating process.

2.2 The role of probability

Starting with a specific relationship between the data and the unknown parameter, M_θ, both Bayesian and frequentist methods attempt to characterize the value of θ that is most in line with a single, observed data set. Although the estimating procedures of both paradigms are the same, the use of probability may differ depending on which statistical principles are prioritized.

Bayesian inference requires one to characterize the unknown parameter probabilistically both before and after observing data. The probability law may be arbitrary (subjective priors) or derived from existing knowledge, theories, and/or data (objective priors). Although the requirement of a probability law to describe uncertainty in the value of θ before observing data is often relaxed in practice (i.e., using "improper priors"), the philosophical justification for inference from a Bayesian methodology cannot permit this violation. For this reason, we restrict our discussion of *probabilistic certainty* to those cases of Bayesian inference that use a proper prior.

https://doi.org/10.1017/psa.2023.174 Published online by Cambridge University Press

The first major conundrum we encounter within a Bayesian paradigm is presented in example 2.1.

Example 2.1. Suppose the second participant in Bayes's billiard table experiment is told that out of their $n = 14$ rolls, ball O has landed to the left of ball W three times. The location of ball W, known only to the first participant, is $\theta_0 = 0.34$ units from one edge of the billiard table. Player 2 decides to use a Bayesian approach and characterizes their uncertainty about θ_0 with a Beta$(0.5,\ 2)$ prior distribution.

Suppose player 2 has a friend who has been watching this game from the beginning. This friend (player 3) may also use a Bayesian approach to assess the location of ball W but may differ in their characterization of uncertainty. Let's say player 3 chooses to model their uncertainty with a Beta$(3,\ 2)$ distribution.

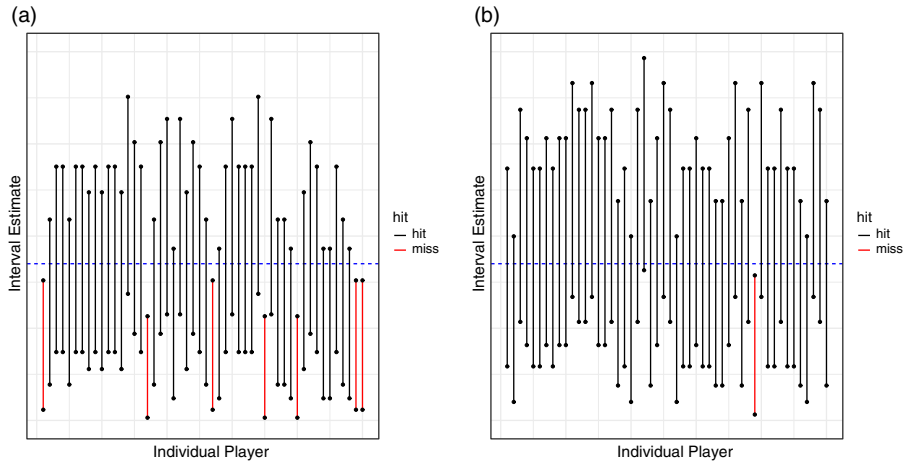
Player 2 is 95% confident that the location of ball W is between 0.058 and 0.432, whereas player 3 is just as confident that the location of ball W is actually between 0.133 and 0.535. (More accurately, but no less confusing, player 2 can logically state that there is a 95% chance that ball W lies between 0.058 and 0.432, whereas player 3 can assert that there is a 95% chance W lies between 0.133 and 0.535.)

In this example, the disagreement between these two intervals is inconsequential. However, in more realistic applications where θ represents, say, the efficacy of a new drug, this disagreement is dire. One interval contains the value 0.5, whereas the other does not. Whose analysis is preferable? Practically, in any situation where M_θ is agreed upon, the basis for any scholarly debate comes down to the choice of prior. Without an appeal to the notion of error control, there is no readily agreed-upon answer to compare posterior distributions resulting from different priors. To this point, another conundrum is presented in example 2.2.

Example 2.2. Revising the billiard table setting, suppose the entire bar has crowded around the trio, intrigued by their bizarre game. Player 1 alone knows the location where ball W stopped ($\theta_0 = 0.34$) but decides to let another 49 members of the captivated audience roll ball O a total of $n = 14$ times and patiently proceeds to tell each person the number of times ball O lands to the left of ball W. By this time, there has been a lengthy discussion on which prior to use for the unknown location of ball W, and now everyone selects a Beta$(0.5, 5)$ prior. Because each of the now 50 players has their own observed data, each player derives their own posterior conditioned on these data. Despite the fact that everyone uses the same prior and the same likelihood, the different data yield different posterior distributions and different inferential conclusions about the location of ball W. Figure 1a shows the resulting 95% credible intervals for all 50 players, who are equally certain in their estimation of the location of ball W. In the end, a handful of players will feel cheated, having missed the target (blue dashed line) despite using the exact same procedure!

The inconsistent conclusions in example 2.2 result from the fact that although the likelihood implies a probabilistic structure governing the behavior of the data (a type of empirical probability), there is no consideration of hypothetical data within a

https://doi.org/10.1017/psa.2023.174 Published online by Cambridge University Press



Seven (red) of the 50 Bayesian credible intervals miss θ_0.

One (red) of the 50 frequentist confidence intervals misses θ_0.

Figure 1. Fifty independent experiments produce 95% credible/confidence interval estimates for θ_0 (blue) in the billiard table experiment.

Bayesian paradigm. This unfortunately results in uncalibrated inferential conclusions. We consider this a serious issue because the implications extend beyond any individual practitioner, instead calling into question the general validity of statistical inference.

A frequentist take on Bayes's theorem understands a prior as an initial distribution estimator for a parameter and the resulting posterior as an updated (or improved) version of this estimator. Estimators, whether distribution, point, or interval, are evaluated according to their performance properties relative to the inferential context. Reid and Cox (2015) call inferential methods that target a standard of performance *calibrated*.

Definition 2.1 "A method of statistical inference is **calibrated** if it does not produce systematically misleading conclusions when used repeatedly" (Reid and Cox 2015).

Figure 1b shows the resulting confidence intervals if the audience instead used a frequentist approach. There is no longer a need for a prior, and this approach is designed to control the number of individuals unable to capture θ_0.

Although the aim of Bayesian methodology is to produce a posterior (i.e., a data-informed distribution estimator for a model parameter), most practitioners agree on the importance of calibrated inferential conclusions. On the other hand, frequentist methods are designed to ensure calibration but do not often target a distribution estimator. Distribution estimators can be much more informative than a single (confidence) interval or *p*-value and empower the researcher by reinforcing the role of statistics as a reliable language of science rather than a prescriptive rule book declaring what constitutes a scientific discovery. In the next section, we present a framework orienting statistical inference around the performance of a calibrated distribution estimator called a *confidence distribution* (CD).

https://doi.org/10.1017/psa.2023.174 Published online by Cambridge University Press

3 Calibrated distribution estimators

Statistical estimation, the primary step toward statistical inference, is motivated by a concrete goal: use observable data to approximate the value of an unknown model feature. Statistical inference is distinguished by the incorporation of probabilistic logic into a statement of estimation. Thus far we have discussed Bayesian and frequentist methods of statistical inference. Here, we present a new framework to directly incorporate probabilistic logic into the estimator itself (much like a posterior distribution) while ensuring calibrated inferential conclusions (as in classical frequentist methodologies).

3.1 Confidence distribution defined

Although the concept of CDs can be most immediately traced back to Fisher (1922) and Cox (1958), the modern interpretation of these functions is a major development within statistical theory (Xie and Singh 2013; Schweder and Hjort 2016). This development untangles CDs from Fisher's flawed fiducial reasoning. Today, CDs enjoy new attention from the statistical community because the modern definition (which follows) avoids the circular reasoning of Fisher's fiducial inference by defining the estimator according to (calibrated) performance standards. This definition is aligned with Xie and Singh (2013) in understanding a CD as an estimator *for* the parameter of interest rather than a distribution *of* the parameter.

Definition 3.1 A sample-dependent function on the parameter space, that is, a function on $\mathcal{X} \times \Theta$, $H_n(\cdot) = H_n(X, \cdot)$, is called a *CD* for $\theta \in \Theta$ if

[R1] For each given sample $x \in \mathcal{X}$, the function $H_n(\cdot) = H_n(x, \cdot)$ is a distribution function on the parameter space Θ; and
[R2] The function can provide confidence intervals for the parameter θ at all confidence levels (Xie and Singh 2013).

In his discussion to Xie and Singh (2013), David Cox states that CDs "provide simple and interpretable summaries of what can reasonably be learned from data." CDs are a general descriptive concept in statistical estimation. Any data-dependent function defining a distribution on the parameter space, whose performance is calibrated across random samples of data, is a CD. Bayesian posteriors can be CDs, as can bootstrap distributions. However, there are many other ways to obtain a CD.

Example 3.1. Suppose our data are random observations from an $N(\mu, \sigma^2)$ distribution where $\sigma^2 = 1$. Suppose further than we observe a sample of size $n = 15$ with mean $\bar{x} = 0.06$. A CD for μ can be $H_n(\mu) = H_n(\mu, \bar{X}) = \Phi\left(\frac{\mu - \bar{X}}{1/\sqrt{15}}\right)$ because $H_n(\mu)$ satisfies the two requirements in definition 3.1, where $\Phi(\cdot)$ is a standard normal (cumulative) distribution function.

Consider two ways in which one could derive the CD in example 3.1. If the primary interest is to assess the evidence for a particular claim about μ, say, C_μ, it makes sense to consider the p-value of this claim, given the observed data. Allowing c to vary

https://doi.org/10.1017/psa.2023.174 Published online by Cambridge University Press

produces a function defined on the parameter space, that is, the CD $H_n(\mu)$. Alternatively, one could derive this distribution estimator as the posterior from an objective Bayesian procedure assuming a flat (improper) prior. Regardless of how the CD is derived, the resulting function defines a probability distribution on the parameter space. Once data are observed, random draws from this distribution can be artificially generated to produce (nonrandom) observed instances of a *CD random variable*.

Definition 3.2 Let $x_{\text{obs}} = (x_1, \ldots, x_n)$ be an observed sample of data drawn from a distribution involving parameter θ, and let $H_n(\theta)$ be a CD for θ. Then, $\theta^*_{\text{CD}}|x_{\text{obs}} \sim H_n(\cdot)$ is referred to as a **CD random variable** (Xie and Singh 2013).

The notion of a CD random variable is distinct from the idea of a random parameter. A CD random variable is a randomized estimator for a fixed unknown, much in the same way that a bootstrap procedure produces a randomized estimator for a parameter. On the left-hand side of equation 1, θ^* can be understood as a CD random variable for θ. Furthermore, a CD, by definition, produces calibrated inferential conclusions about θ. Thus, a CD and a CD random variable provide measures of corroboration for any inferential statement about θ while embodying the duality of a parameter.

3.2 Severity and confidence distributions

We now explore a connection between CDs and severe inference whereby error probabilities distinctly characterize the severity with which a claim about θ has been tested (Mayo 2018). Because a CD is constructed with respect to error probabilities, it also contains the mathematical information necessary to quantify the severity of any related test of a claim about θ.

Definition 3.3 Strong severity requirement. "We have evidence for a claim C just to the extent it survives a stringent scrutiny. If C passes a test that was highly capable of finding flaws or discrepancies from C, and yet none or few are found, then the passing result ... is evidence for C" (Mayo 2018, 14).

Continuing with example 3.1, as a functional estimator, $H_n(\mu)$ contains a wealth of inferential information about the unknown parameter μ. For instance, one may calculate the p-value associated with testing claim $C_\mu : \mu \leq c$ by evaluating $\int_{-\infty}^{c} dH_n(\mu) = H_n(c)$. The p-value is a mathematical summary of the support our estimator, $H_n(\mu)$, lends to claim C_μ. This is a useful summary based on the pre-data rationale to use a method that controls error probabilities. Post-data, we want to understand how severely our claim has been tested. In this example, the severity with which claim C_μ is tested is inversely related to $H_n(c)$. Figure 2 shows $H_n(c)$ (black) and the severity curve (blue) for varying values of c. The similarities between figure 2 and figure 3.3 of Mayo (2018) are not coincidental and indeed illustrate the relationship between severity and CDs.

As an estimator, a CD is not dependent on any particular hypothesis but summarizes the inferential information about θ contained in the data. A post hoc severity assessment of this information can always be calculated because the severity will be a function of a CD. It seems that, provided the strong severity requirement is met, a CD may supply not only an empirical but also an evidential interpretation of the conclusions of a test about θ. Therefore, we wonder, is a prior a requirement for

https://doi.org/10.1017/psa.2023.174 Published online by Cambridge University Press

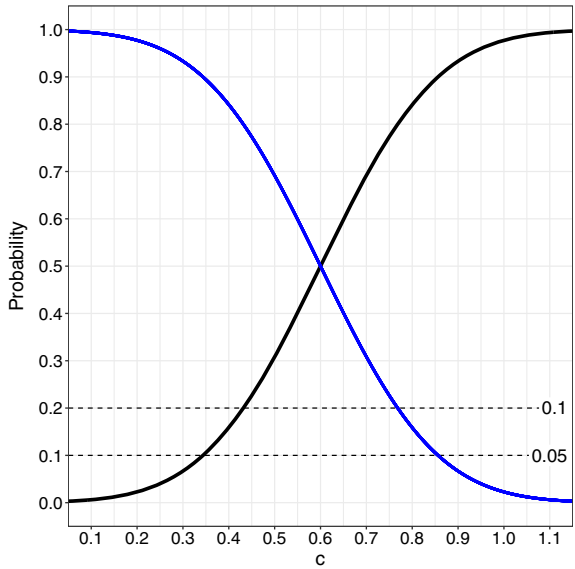


Figure 2. For example 3.1, where $\bar{x}_{\text{obs}} = 0.6$, a CD for μ (black) is related to the severity (blue) of testing the claim $C_\mu : \mu \leq c$. Significance levels at $\alpha = 0.1$ and $\alpha = 0.05$ are marked with dashed lines.

characterizing epistemic uncertainty probabilistically? And if so, under what conditions on the prior?

4 Conclusion

A duality in model parameters arises as an unlikely connection between frequentist and Bayesian inference tasks. Statistical inference through CDs codifies this duality while connecting the developments of error probability and severity. In this context, Neyman's "confidence" is a corroboration measure (similar to the measure to control erroneous conclusions in the severity development). In contrast to a posterior, a CD need not be interpreted as a probabilistic statement about the parameter; rather, it is a data-dependent estimator for which a particular behavioral property holds. A CD contains inferential information corresponding to any level of confidence or significance as a calibrated method. CDs can be derived through Bayesian or frequentist procedures, guaranteeing calibrated performance regardless. It is not necessary to assume a prior to derive a CD that characterizes the most plausible values of an unknown parameter. This begs the philosophical question, If a (perhaps entirely subjective) prior in Bayesian inference can have epistemic meaning, what about a data-informed, calibrated distribution estimator?

We briefly mentioned a third (less popular) inferential framework known as *fiducial inference* (Fisher 1922). Statisticians today understand fiducial inferential procedures as *inversion methods* that solve a structural (algorithmic) model for the target parameter (Hannig et al. 2016). The notion of parameter duality can be found in classical fiducial literature; however, Fisher's original interpretation is incoherent

https://doi.org/10.1017/psa.2023.174 Published online by Cambridge University Press

and leads to a number of paradoxes. Instead, we understand modern fiducial approaches as another procedure whereby one may obtain a CD.

A calibrated statistical inferential paradigm afforded by distribution estimators bypasses the historically conflicting perspectives on parameters and probabilities. In fact, there are many equivalent calibrated distribution estimators besides CDs. We find the concept of CDs appealing because of the simplicity of the required conditions. A CD has a performance guarantee (unlike subjective Bayesian posteriors), and as an estimator, there are mathematical ways to consider optimal choices of CDs for any given problem. Objective Bayesian procedures often yield a posterior that ends up being a CD, but the definition of a CD lends procedural guidance as to how to choose among different posteriors in addition to permitting alternate approaches (besides Bayes's theorem) to derive a CD.

References

Bayes, Thomas. 1764. "An Essay toward Solving a Problem in the Doctrine of Chances." *Philosophical Transactions of the Royal Society of London* 53:370–418.

Beaumont, Mark A. 2019. "Approximate Bayesian Computation." *Annual Review of Statistical Applications* 6 (1):379–403. doi: 10.1146/annurev-statistics-030718-105212.

Berger, James O., Jose M. Bernardo, and Dongchu Sun. 2015. "Overall Objective Priors." *Bayesian Analysis* 10 (1):189–221. doi: 10.1214/14-BA915.

Cox, David R. 1958. "Some Problems Connected with Statistical Inference." *Annals of Mathematical Statistics* 29 (2):357–72.

Efron, Bradley. 2000. "The Bootstrap and Modern Statistics." *Journal of the American Statistical Association* 95 (452):1293–96. doi: 10.2307/2669773.

Fisher, Ronald A. 1922. "On the Mathematical Foundations of Theoretical Statistics." *Philosophical Transactions of the Royal Society of London, A* 222 (594–604):309–68.

Hannig, Jan, Hari Iyer, Randy C. S. Lai, and Thomas C. M. Lee. 2016. "Generalized Fiducial Inference: A Review and New Results." *Journal of the American Statistical Association* 111 (515):1346–61. doi: 10.1080/01621459.2016.1165102.

Mayo, Deborah G. 2018. *Statistical Inference as Severe Testing: How to Get beyond the Statistics Wars*. Cambridge: Cambridge University Press.

Reid, Nancy, and David R. Cox. 2015. "On Some Principles of Statistical Inference." *International Statistics Review* 83 (2):293–308. doi: 10.1111/insr.12067.

Schweder, Tore, and Nils Lid Hjort. 2016. *Confidence, Likelihood, Probability: Statistical Inference with Confidence Distributions (Cambridge Series in Statistical and Probabilistic Mathematics, Series Number 41)*. Cambridge: Cambridge University Press.

Thornton, Suzanne, and Minge Xie. Forthcoming. "Bridging Bayesian, Frequentist, and Fiducial Inferences Using Confidence Distributions." In *Handbook of Bayesian, Fiducial, and Frequentist Inference*, edited by James Berger, Xiao-Li Meng, Nancy Reid, and Minge Xie. New York: Chapman & Hall.

Xie, Minge, and Kesar Singh. 2013. "Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review (Including Discussions and Rejoinder)." *International Statistics Review* 81 (1):3–77. doi: 10.1111/insr.12000.

Cite this article: Thornton, Suzanne and Minge Xie. 2024. "An Exploration of Parameter Duality in Statistical Inference." *Philosophy of Science*. https://doi.org/10.1017/psa.2023.174