Pak. J. Statist.
200x, Vol. xx(x), xx-xx

STATISTICAL INFERENCE WITH A SINGLE OBSERVATION OF $N(\theta, 1)$

Duncan Ermini Leaf[1], **Jun Hui**[2], and **Chuanhai Liu**[3]

[1,3] Department of Statistics, Purdue University, West Lafayette, USA
[2] Mathematics Department, Heifei University of Technology, Heifei, China
Email: [1] dleaf@stat.purdue.edu [2] jhui@stat.purdue.edu
[3] chuanhai@stat.purdue.edu

ABSTRACT

We consider some fundamental issues in statistical inference by focusing on inference about the unknown mean θ of the Gaussian model $N(\theta, 1)$ with unit variance from a single observed data point X. A closer look at this seemingly simple inference problem reveals a limitation of objective Bayesian posteriors in that they cannot be interpreted as valid posteriors when combining certain types of information. A new solution to inference about θ from X is proposed. The proposed method is based on the fiducial distribution of θ given X, but with a new Weak Belief rule of combination for constraint-type information. It is shown that the proposed approach is promising for constrained statistical inference.

KEYWORDS

Constrained statistical inference, Credible inference, Dempster-Shafer theory, Fiducial argument, Lower and upper probabilities, Maximal belief

2000 Mathematics Subject Classification: 62F30

1 INTRODUCTION

A Silver Jubilee celebration seems a fitting time for statisticians to consider the past and future of our field. In this article we focus on the course of probabilistic reasoning toward statistical inference. Probabilistic reasoning requires a meaningful interpretation of probabilities that can be understood by applied statisticians and scientists. To avoid unnecessary philosophical debates, the use of probability in this paper is interpreted in terms of frequency or proportion, as is typically understood and used by practitioners in building sampling models for data. When personal probability is involved, we shall accept its use so long as its numerical values are calibrated to have a sensible frequency interpretation. In what follows, the phrases "long-run frequency" and "repeated experiments" are used to imply that the corresponding probability is interpreted in terms of frequency. Consequently, the needed mathematical tool is the standard theory of additive probability which satisfies the Kolmogorov axioms.

Let $F(x|\theta)$ denote the cdf of the sampling distribution for the observed data X in its sample space $\mathcal{X}$ with parameter θ in the parameter space Θ. Denote by $f(x|\theta)$ the pdf of $F(x|\theta)$. To address some fundamental issues in statistical inference, this paper focuses on a simple Gaussian example. The sampling model for the observed data point X is the univariate normal $X \sim N(\theta,1)$ with unit variance and unknown mean θ in $\Theta = \mathcal{R} = (-\infty,\infty)$. The cdf and pdf of the sampling distribution are then

$$F(x|\theta) = \Phi(x-\theta) \qquad (x \in \mathcal{X} = \mathcal{R}, \theta \in \mathcal{R}) \tag{1.1.1}$$

and

$$f(x|\theta) = \phi(x-\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} \qquad (x \in \mathcal{X} = \mathcal{R}, \theta \in \mathcal{R}), \tag{1.1.2}$$

where $\Phi(.)$ and $\phi(.)$ denote the cdf and pdf of the standard normal distribution. This paper focuses on inference about θ from the observed point X *without* prior information on θ.

From a problem-solving perspective any logically sound approach to inference about θ ought to proceed with a good understanding of assumptions about the randomness in the postulated sampling model for the observed data point X. Statistical inference is then reduced to propagating the uncertainties specified in the sampling model to inference about θ. We call this approach the fundamental principle of inferential problem solving.

That said, the authors see Fisher's fiducial argument heading in such a sensible direction, although it remains to be fully and correctly developed (see, *e.g.*, Zabell, 1992; Zhang & Liu, 2009). A fiducial solution to inference about θ in absence of prior knowledge about θ is discussed in Section 3, where the fiducial distribution is (*e.g.*, Fisher, 1959; Dawid & Stone, 1982):

$$\theta|X \sim N(X,1) \qquad (\theta \in \mathcal{R}). \tag{1.1.3}$$

Applying Bayes' theorem (see Section 2) with the non-informative prior $\pi(\theta) \propto 1$ for θ, without questioning the meaning of Bayes' theorem, gives the posterior $N(X,1)$ for inference about θ, *i.e.*,

$$\pi(\theta|X) = \phi(\theta - X) \qquad (\theta \in R), \tag{1.1.4}$$

which is the same as the fiducial distribution (1.1.3). This has been known as objective Bayesian inference. One could argue that the posterior distribution (1.1.4) makes sense because, computationally, it reproduces the fiducial posterior (1.1.3). In this regard, the results of Lindley (1958) can be used to identify a class of single parameter models for which there exist Bayesian distributions that exactly reproduce the fiducial distributions. As a result, objective Bayes and the fiducial approach share favorable properties, such as data driven analysis, as well as the undesirable property that posteriors should be interpreted differently from valid (non-"objective") Bayesian posteriors; see Section 2. Most importantly, care must be taken when combining certain types of information into fiducial and objective Bayesian posteriors. It is shown in Section 2 that, compared to valid Bayesian inference in the context of situation-specific inference, objective Bayesian posteriors cannot be interpreted as valid posteriors when used for combining constraint-type information.

The remaining sections of this paper are organized as follows: Section 2 reviews two fundamental properties of valid Bayesian posteriors for statistical inference and shows that the objective Bayesian posterior (1.1.4) is problematic when combined with additional information that is given in the form of $a \leq \theta \leq b$. Section 3 presents a new solution to inference about θ from X with some types of constraints, including (i) $a \leq \theta \leq b$ and (ii) $\theta \in \{a, b\}$, where $a < b$. The approach we take follows the Maximal Belief method of Zhang and Liu (2009). Section 4 concludes with a few remarks, which draw attention to fundamental issues in statistical inference.

2 BAYESIAN INFERENCE

2.1 Valid Bayesian inference

The authors define a *valid* prior distribution $\pi(\theta)$ $(\theta \in \Theta)$ as one that represents the correct sampling distribution for θ, and thereby, the hyper-population $\{F(\cdot|\theta)\}$ in repeated experiments. A valid prior distribution can also be personal, but in this case the authors expect the numerical values of personal probability to be consistent with frequency probability so that the standard probability theory can be applied. With a valid prior distribution

$\pi(\theta)$, where $\int_\Theta \pi(\theta)d\theta = 1$, the conditional distribution of θ given the observed data X,

$$\pi(\theta|X) = \frac{\pi(\theta)f(X|\theta)}{\int_\Theta \pi(\theta)f(X|\theta)d\theta}, \tag{2.2.1}$$

is given by Bayes' theorem. When $\pi(\theta)$ is a valid prior distribution, $\pi(\theta|X)$ is called a valid posterior distribution.

To construct an example of a valid posterior that has the same form as the objective Bayesian posterior (1.1.4), we take

$$\theta \sim N\left(0, \tfrac{c}{c-1}\right) \quad \text{and} \quad Y|\theta \sim N(\theta, c), \tag{2.2.2}$$

where $Y = cX$ and $c > 1$. It follows from (2.2.1) that the valid posterior is $\theta|X \sim N(X,1)$, *i.e.*,

$$\pi(\theta|X) = \phi(\theta - X) \qquad (\theta \in R), \tag{2.2.3}$$

a familiar Gaussian distribution whose pdf and cdf are displayed in Fig. 1.

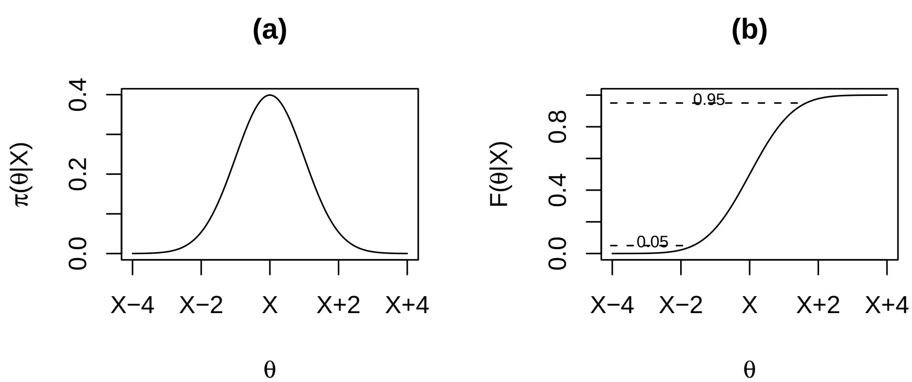

Figure 1: The pdf (a) and cdf (b) of the posterior $\theta|X \sim N(X,1)$ obtained from the model (2.2.2)

For conceptual clarity, denote by θ^* the unobserved realization of θ in the specific experiment where X was collected. Suppose that the observed value of X is 0.5 and that we are interested in the "sharp" or point assertion that $\theta^* = 0$, denoted by $A = \{\theta^* : \theta^* = 0\}$. When the posterior (2.2.3) is taken as a Dempster-Shafer (DS; Dempster, 2008) model

(DSM) for inference about θ^*, the DS output for this assertion is $(p,q,r) = (0,1,0)$, with p as the (personal) probability for the truth of A, q as the probability against the truth of A, and $r = 1-(p+q)$ as the remaining probability, known as the probability of "don't know." By using (2.2.3) as a DSM, our probability against $\theta^* = 0$ (or any other point assertion, for that matter) is always one. Obviously, statisticians would not feel comfortable using such a (p,q,r) output for inferring the truth or falsity of A. Typically, applied statisticians interpret the posterior as that (*i*) it is unlikely that θ^* lies outside of $[X-2, X+2]$ and (*ii*) it is very likely that θ^* is in the interval $[X-2, X+2]$, but it is difficult to claim with confidence that θ^* is very close to a particular point located in this interval. This interpretation is necessarily personal. Moreover, this line of thought implies an informal inferential probability model representing degrees of belief about the realization of θ^*. Martin, Zhang, and Liu (2009) extended this idea and proposed to specify a DSM that precisely represents the belief so that meaningful DS (p,q,r) output for assertions can be calculated formally.

2.2 Situation-specific inference

We call the above inference about θ^* from the observed data X *situation-specific*. The term "situation-specific" has usually been taken to mean "conditioned on data". This term or, for clarity, "strong situation-specific" is used here to emphasize that inference is both "conditioned on data" and "realization-specific" in the Bayesian context; see Martin et al. (2009) for more discussion. The concept of "situation-specific" makes it possible to bring both Bayesian and Frequentist schools of thought into a unified framework of statistical inference. That is, solving inferential problems on a specific value no matter whether it is an unknown fixed quantity or an unknown realization from a known distribution.

In the present context of making inference about θ from a single observation $Y = cX \sim N(\theta, c)$ with the prior distribution $\theta \sim N\left(0, \frac{c}{c-1}\right)$, it is obvious that we would not use the posterior distribution $\theta|X \sim N(X,1)$ as a DSM to represent our personal uncertainty about the particular realization θ^*. As argued earlier, statisticians and scientists make use of intervals based on $\theta|X \sim N(X,1)$ to represent their uncertainty about θ^*. Intervals of this kind are personal and take different forms in practice for different assertions of interest. Following Zhang and Liu (2009), we use random intervals and their corresponding DSMs for a systematic treatment. In particular, we consider three types of random intervals:

1. Left-sided

$$L(\theta) = \{\theta^* : \theta^* \leq \theta\} = (-\infty, \theta], \tag{2.2.4}$$

2. Right-sided

$$R(\theta) = \{\theta^* : \theta^* \geq \theta\} = [\theta, \infty), \tag{2.2.5}$$

   and

3. Centered

$$C(\theta) = \{\theta^* : |\theta^* - X| \leq |\theta - X|\} = X \pm |\theta - X|, \tag{2.2.6}$$

where $\theta|X \sim N(X,1)$.

The attractive feature of using DSMs to represent our uncertainty about θ^* given the observed data X is that we can compute the induced beliefs on the truth or falsity of any assertion. Consider, for example, the DSM specified by the random interval $L(\theta)$ and the assertion $A = \{\theta^* : \theta^* = 0\}$. The probability that the random interval $L(\theta)$, believed to contain θ^*, does not contain 0 is the implied belief against the truth of A. Using this DSM, we have the following (p,q,r) for A:

$$p(A) = 0,\ q(A) = \Pr(\theta < 0|X) = \Phi(-X),\ \text{and } r(A) = \Phi(X).$$

Similarly, if we choose the DSM specified by the random interval $C(\theta)$, then the (p,q,r) for A becomes

$$\begin{aligned}&p(A) = 0,\ q(A) = \Pr(|X| > |\theta - X||X) = 2\Phi(|X|) - 1,\ \text{and}\\ &r(A) = 2 - 2\Phi(|X|).\end{aligned}$$

2.3 Two characteristics of valid posteriors

Here we consider properties of valid Bayesian posteriors for inference. In particular, the long-run frequency distributions of

$$Q_L(X,\theta^*) = \Pr(\theta^* \notin L(\theta)|X) \tag{2.2.7}$$

and

$$Q_R(X,\theta^*) = \Pr(\theta^* \notin R(\theta)|X), \tag{2.2.8}$$

corresponding to familiar tail probabilities, are of interest. We refer to the resulting inferential property, formally summarized into the following theorem, as the *situation-specific* property.

Theorem 1. *(Situation-specific) Suppose that* $\theta \sim N\left(0, \frac{c}{c-1}\right)$ *and* $(cX)|\theta \sim N(\theta, c)$ *with* $c > 1$*. Then* $Q_L(X, \theta^*) \sim Unif(0,1)$ *and* $Q_R(X, \theta^*) \sim Unif(0,1)$.

Proof. Note that according to the definition (2.2.7), $Q_L(X, \theta^*)$ is the cdf of the conditional distribution of $\theta|X$ evaluated at θ^*. Since this cdf is continuous and θ^* is a random sample from this distribution in repeated experiments, it holds that $Q_L(X, \theta^*)$ is distributed uniformly on the interval from zero to one. Similarly, it can be shown that $Q_R(X, \theta^*) \sim \text{Unif}(0,1)$. ∎

Theorem 1 implies that the DS (p, q, r) output based on either random set, $L(\theta)$ or $R(\theta)$, has desirable frequency properties. Another important property, concerning inference with constraints on parameters, is on combining the information given by the posterior $\theta|X \sim N(X, 1)$ and an additional piece of information of the form

$$\theta \in [a, b] \qquad (-\infty \le a < b \le \infty), \tag{2.2.9}$$

where a and b are known constants. Probability theory defines the rule of combination that results in the combined posterior,

$$\pi(\theta|X, a \le \theta \le b) \propto \phi(\theta - X) \qquad (\theta \in [a, b]). \tag{2.2.10}$$

For this combined posterior, the corresponding situation-specific property (Theorem 1) holds. Formally, we have the following theorem, a property of what we call Bayes' rule of combination, which can be viewed as Dempster's rule of combination (Dempster, 2008) when the Bayesian posterior is taken as a DSM for inference about θ^*.

Theorem 2. *(Combining information) Assume that the assumptions of Theorem 1 hold and, in addition, that* θ^* *satisfies the constraint (2.2.9). Then* $Q_L(X, \theta^*) \sim Unif(0,1)$ *and* $Q_R(X, \theta^*) \sim Unif(0,1)$*, where* $Q_L(X, \theta^*)$ *and* $Q_R(X, \theta^*)$ *are given in (2.2.7) and (2.2.8) but their corresponding probabilities determined by (2.2.10).*

Proof. Note that the constrained posterior (2.2.10) is continuous. The proof is similar to that of Theorem 1. ∎

2.4 A difficulty in interpreting objective Bayesian posteriors

We now turn to objective Bayesian inference about θ^* from a single observation X of $N(\theta, 1)$ with no prior knowledge about θ. The objective Bayesian posterior $\theta|X \sim N(X, 1)$ is obtained from Bayes' theorem by taking the so-called non-informative prior $\pi(\theta) \propto 1$. It has been well demonstrated that this posterior has nice frequency properties when used for situation-specific inference. This is due to the following results that are similar to those of Theorem 1 with a slightly different proof.

Theorem 3. *(Situation-specific) Suppose that $X|\theta \sim N(\theta,1)$ with no prior knowledge about θ. Then $Q_L(X,\theta^*) \sim Unif(0,1)$ and $Q_R(X,\theta^*) \sim Unif(0,1)$, where $Q_L(X,\theta^*)$ and $Q_R(X,\theta^*)$ are defined in(2.2.7) and (2.2.8) but with $\theta|X \sim N(X,1)$, obtained by applying Bayes' theorem with the non-informative prior $\pi(\theta) \propto 1$.*

Proof. It follows from (2.2.7) that for the given pair (X,θ^*) in an experiment

$$Q_L(X,\theta^*) = \Pr_{\text{obj}}(\theta \le \theta^*|X) = \Phi(\theta^* - X),$$

where the "obj" subscript in $\Pr_{\text{obj}}$ indicates that the corresponding probability calculation is with respect to the objective Bayesian posterior $\theta|X \sim N(X,1)$. Thus, in repeated experiments that require X to be the random variable, $\Phi(\theta^* - X)$, and thereby, $Q_L(X,\theta^*)$, follows the Unif$(0,1)$ for any θ^*. Similarly, it can be shown that $Q_R(X,\theta^*) \sim \text{Unif}(0,1)$. ∎

Theorem 3 makes objective Bayes attractive for inference without prior information about θ. However, the objective Bayesian posterior does not have the desirable property corresponding to Theorem 2 when constraints are introduced. Suppose that in addition to $X|\theta \sim N(\theta,1)$, it is known that (2.2.9) holds for some fixed pair of a and b. In this case, the combined objective Bayesian posterior has the cdf,

$$F(\theta|X, a \le \theta \le b) = \frac{\Phi(\theta - X) - \Phi(a - X)}{\Phi(b - X) - \Phi(a - X)}. \tag{2.2.11}$$

For the results corresponding to Theorem 2 to hold, $F(\theta^*|X, a \le \theta \le b)$ must follow the Unif$(0,1)$ distribution for fixed a and b in repeated experiments with $X - \theta^* \sim N(0,1)$ and no prior knowledge about θ^*. This is not true. For example, in the case of $\theta^* = a$ in repeated experiments $F(\theta|X, a \le \theta \le b) = 0$ with probability one and in the case of $\theta^* = b$ in repeated experiments $F(\theta|X, a \le \theta \le b) = 1$ with probability one. The lack of results corresponding to Theorem 2 leads to the conclusion that *in general, objective Bayesian posteriors cannot be interpreted or used for statistical inference in the same way as those obtained with valid priors*. More discussion on objective Bayes is given in Section 4.

3 SITUATION-SPECIFIC FIDUCIAL INFERENCE USING WEAK BELIEFS

In the case with no prior information about θ, all the known information is given by the postulated sampling model $X|\theta \sim N(\theta, 1)$ for the single observation X. The uncertainty is entirely due to the randomness of

$$Z = X - \theta \qquad (Z \sim N(0,1),\ \theta \in R\ ). \tag{3.3.1}$$

The unknown value of θ in a particular experiment, denoted by θ^*, is uniquely determined by the realization of Z^* in that experiment, where $X = \theta^* + Z^*$. However, having only observed X we can only say that the point (θ^*, Z^*) falls somewhere on the line $X = \theta + Z$. The problem of inference is then to represent our uncertainty about Z^* from the observed data X, equation (3.3.1), and the fact that $Z \sim N(0,1)$. This approach leads to direct probabilistic inference whereas objective Bayesian inference can be viewed as indirect inference.

In the context of the current paper's focus on the long-run frequency interpretation of probability, Fisher's fiducial distribution (*e.g.*, Fisher, 1959), to be used for inference about θ from X has the cdf,

$$F(\theta|X) = \Phi(\theta - X). \tag{3.3.2}$$

This distribution can be viewed as obtained from (3.3.1) by believing that under complete ignorance about θ or, more exactly, θ^*, the knowledge of (3.3.1) provides no evidence to alter the long-run frequency distribution of Z. That is, $Z|X \sim N(0,1)$. In this case, equation (3.3.1) serves as a transformation to propagate our uncertainty about Z^* to θ^* in the particular experiment with observed data X. Then, the fiducial distribution (3.3.2) can be used as a DSM for making inference about θ^* (Zhang & Liu, 2009).

Note that the fiducial cdf $F(\theta|X)$ given in (3.3.2) is identical to the objective Bayesian posterior, but with a well defined interpretation of the underlying probability. Hence, Theorem 3 for objective Bayes applies to the fiducial distribution as well. It is the authors' opinion that objective Bayes and fiducial inference both aim to accomplish the same task, with fiducial inference being the more direct method, but whose ultimate acceptable version has yet to be developed. Consequently, fiducial distributions also share with objective Bayesian posteriors the difficulty in inference with constraints on unknown parameters.

The above discussion motivates a way of resolving the problem regarding objective Bayes in Section 2.4 by taking a closer look at inference with constraints in the light of the

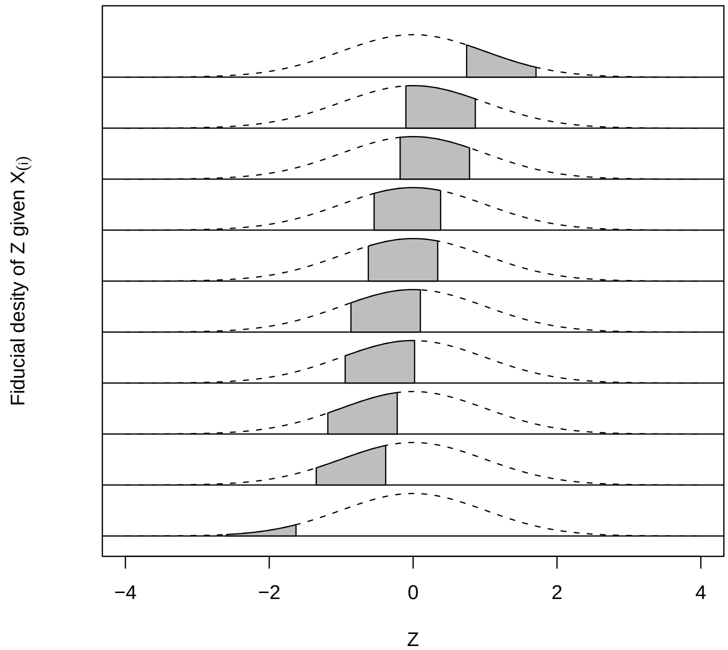

Figure 2: Illustration of constraints on Z from observations of X in repeated experiments.

fiducial setting. Recall that the constraint $\theta \in [a, b]$ is considered in the simple Gaussian example, $\theta | X \sim N(\theta, 1)$. From (3.3.1), this constraint implies that

$$X - b \leq Z^* \leq X - a. \tag{3.3.3}$$

It is seen from (3.3.3) that the usual Bayes' rule of combination is valid for a long-run frequency interpretation for all repeated experiments with a *common* X because the realizations of Z are known to be in the fixed interval (3.3.3). This assumption is obviously not practical because the observed values of X vary in repeated experiments. In this more sensible setting, the constraints for Z vary from experiment to experiment. That is, no common constraint is available for Z across all realizations of X. This phenomenon is illustrated by Fig. 2.

Recall that our goal of inference is to specify our uncertainty about θ^* or, more formally, to specify a DSM for producing (p,q,r) for given assertions of interest. To accomplish this, we consider a predictive random set (PRS) $S(Z)$ for Z^*. To be more specific, take

$$S(Z) = \{Z^* : \ |Z^*| \leq |Z|\} \qquad (Z|X \sim N(0,1)), \tag{3.3.4}$$

which defines a DSM (see Zhang and Liu, 2009). It is easy to prove that

$$\Pr(Z^* \notin S(Z)|Z^*) \sim \text{Unif}(0,1) \tag{3.3.5}$$

as $Z^* \sim N(0,1)$ in repeated experiments. That is, according to Zhang and Liu (2009), $S(Z)$ is credible for inference about Z^*. Note that this PRS always contains the point $Z^* = 0$ and treats $Z^* = -\infty$ and $Z^* = \infty$ as extremal points, meaning that it will be surprising to have values of Z^* far away from zero.

It was shown in Section 2.4 that combining the constraint information (3.3.3) directly with $Z|X \sim N(0,1)$ is problematic when the long-run frequency interpretation of probability is of interest. This raises the following question: what property does Dempster's rule of combination have when applied to the DSM (3.3.4) and (3.3.3)? The DSM (3.3.4) represents a so-called credible belief function for inference about Z^*. In a certain sense, (3.3.3) helps to refine the DSM (3.3.4) for inference about Z^*. However, we may not expect to have a sensible long-run frequency interpretation of the resulting DSM because conflict cases,

$$S(Z) \cap \{Z^* : \ X - b \leq Z^* \leq X - a\} = \emptyset, \tag{3.3.6}$$

exist and can alter the long-run frequency in an uncontrollable way. One way to avoid conflict cases is to weaken the DSM (3.3.4) by minimally enlarging the random set $S(Z)$ so that the resulting intersection with $\{Z^* : \ X - b \leq Z^* \leq X - a\}$ will not be empty. More specifically, we introduce a Weak Belief rule of combination for the present problem (a more general definition will be considered elsewhere),

$$S(Z) \Cap C_X = \{c \in C_X : \min_{s \in S(Z)} |c - s| = \min_{t \in C_X} \min_{s \in S(Z)} |t - s|\}, \tag{3.3.7}$$

where C_X is the set representing the constraints on Z^* for an observed X. When $S(Z) \cap C_X \neq \emptyset$, then $S(Z) \Cap C_X$ is simply the intersection of $S(Z)$ with C_X. In cases where $S(Z) \cap C_X = \emptyset$, then $S(Z) \Cap C_X$ is the set of points in C_X that are closest to the points in $S(Z)$. In the present example, $C_X = [X - b, X - a]$ and we replace

$$S(Z) \cap C_X$$

with

$$S(Z) \Cap C_X = \begin{cases} S(Z) \cap C_X, & \text{if } S(Z) \cap C_X \neq \emptyset; \\ \{X-b\}, & \text{if } |Z| < X-b; \\ \{X-a\}, & \text{if } -|Z| > X-a; \end{cases}$$

to represent our belief about Z^*. An argument for this rule of combination is that the DSM (3.3.4) represents our personal belief and thus can be subject to adjustments or refinements. The combined DSM obtained by this new rule has the following appealing property.

Theorem 4. *Suppose that* $X = \theta^* + Z^*$ *with* $Z^* \sim N(0,1)$ *and given an observation* X*,* $Z^* \in C_X$*. Then the DSM,* $S(Z) \Cap C_X$*, defined in (3.3.7), is credible for inference about* Z^**.*

Proof. It is said that a random set S is credible for inference about a realization Z^* of the random variable Z *iff* the random variable,

$$\Pr\left(S \not\ni Z^* | Z^*\right) \qquad (S \in \mathbb{S} \subseteq 2^{\mathbb{Z}}),$$

as a function of Z^* is stochastically not greater than a uniform random variable on $[0,1]$. Let $q(Z^*) = \Pr\left(S(Z) \not\ni Z^* | Z^*\right)$. It is easy to show that $S(Z)$ is credible for inference about Z^*. That is,

$$q(Z^*) \stackrel{S}{=} \text{Unif}\left(0,1\right),$$

where $\stackrel{S}{=}$ denotes stochastic equality. Let $q_c(Z^*) = \Pr\left(S(Z) \Cap C_X \not\ni Z^* | Z^*\right)$. Note that for any $Z^* \in C_X$,

$$q(Z^*) = \Pr\left(S(Z) \not\ni Z^*, S(Z) \cap C_X \neq \emptyset | Z^*\right) + \Pr\left(S(Z) \cap C_X = \emptyset | Z^*\right)$$

and

$$\begin{aligned} q_c(Z^*) = &\Pr\left(S(Z) \not\ni Z^*, S(Z) \cap C_X \neq \emptyset | Z^*\right) \\ &+ \Pr\left(S(Z) \Cap C_X \not\ni Z^*, S(Z) \cap C_X = \emptyset | Z^*\right). \end{aligned}$$

By monotonicity of the probability measure on Z,

$$\Pr\left(S(Z) \Cap C_X \not\ni Z^*, S(Z) \cap C_X = \emptyset | Z^*\right) \leq \Pr\left(S(Z) \cap C_X = \emptyset | Z^*\right).$$

Therefore,

$$q_c(Z^*) \leq q(Z^*)$$

for all $Z^* \in C_X$ over all realizations of X. It follows that

$$q_c(Z^*) \overset{S}{\leq} \text{Unif}\,(0,1)\,.$$

This completes the proof of the theorem. ∎

For an illustration of Theorem 4, take $a = 0$ and $b = \infty$, *i.e.*, $\theta^* \geq 0$ in repeated experiments. Fig. 3 shows the cdfs of the two end points of the combined random set for inference about θ^* from the observed data X with θ^* constrained to be non-negative. The results show that these cdfs can be mixed continuous and discrete distributions with non-zero point masses at the end points.

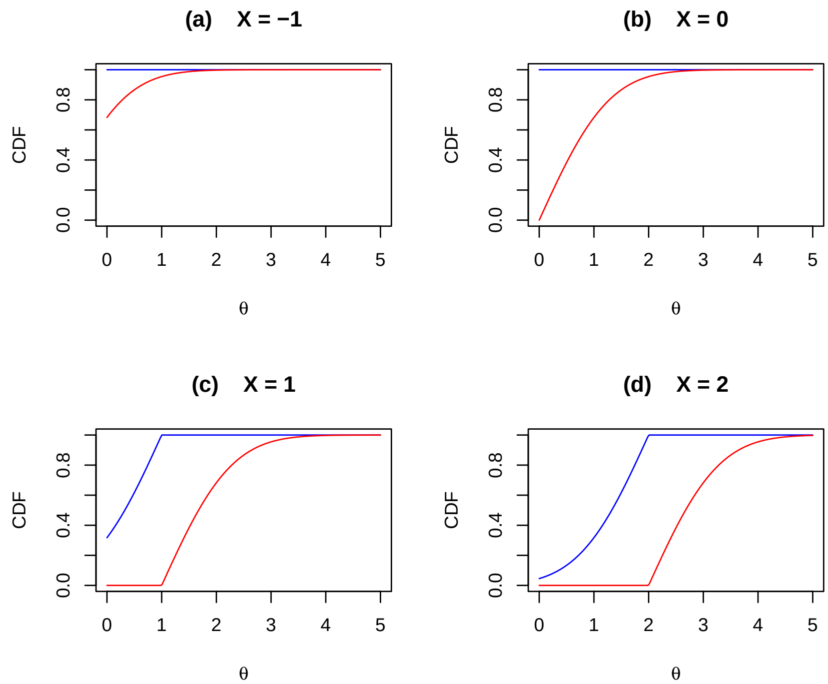

Figure 3: Illustration of conditional cdf for random set with combined information in Theorem 4.

To end this section, we apply the Weak Belief rule of combination (3.3.7) to inference about θ based on a single observation X from $N(\theta,1)$ with the constraint $\theta \in \{a,b\}$, *i.e.*, the parameter space consists of two points a and b, where $a < b$. Here, $C_X = \{X-b, X-a\}$ and the resulting DSM is

$$S(Z) \Cap C_X = \begin{cases} \{X-b\}, & \text{if } |Z| < X - \frac{a+b}{2}; \\ \{X-a\}, & \text{if } -|Z| > X - \frac{a+b}{2}; \\ \{X-b, X-a\}, & \text{if } X = \frac{a+b}{2} \text{ or } \{X-b, X-a\} \subset S(Z). \end{cases}$$

The results are given in Table 1. It is straightforward to show that the probabilities are credible in the sense of Zhang and Liu (2009). For example, under the truth of $\theta^* = a$, the probability *against* the assertion $\theta^* = a$ is

$$\begin{aligned} & \Pr\left(X > \frac{a+b}{2}, 2\Phi\left(X - \frac{a+b}{2}\right) - 1 \geq 1 - \alpha\right) \\ = \ & \Pr\left(X - a > \frac{b-a}{2}, X - a \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{b-a}{2}\right) \\ = \ & \Pr\left(X - a \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{b-a}{2}\right) \\ < \ & \Pr\left(Z \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = \frac{\alpha}{2} < \alpha, \ \text{ where } Z \sim N(0,1). \end{aligned}$$

Table 1: Probability for $\theta^* = a$ or $\theta^* = b$ $(a < b)$ given a single observation X from $N(\theta,1)$, where $\theta \in \{a,b\}$.

Assertion	Probability for the assertion, given X	
	$X \leq \frac{a+b}{2}$	$X > \frac{a+b}{2}$
$\theta^* = a$	$2\Phi\left(\frac{a+b}{2} - X\right) - 1$	0
$\theta^* = b$	0	$2\Phi\left(X - \frac{a+b}{2}\right) - 1$

4 CONCLUDING REMARKS

In this paper, we took a closer look at inference about θ in $N(\theta,1)$ from a single data point X. If the frequency interpretation of probability is taken, several issues arise when incorporating constraint-type information into existing inference methods.

When a valid prior exists, random intervals can be used to form a DSM that allows for credible, situation-specific inference with constraints. However, for many inference problems, it may be difficult to find a sensible prior. Neyman's concept of a confidence interval is appealing to most statisticians and it is helpful to study large sample theory based on Fisher's concepts of consistency, efficiency, and sufficiency (Fisher, 1922). However, an ultimate satisfactory solution should allow us to make direct probabilistic inference with even the smallest sample size. Objective Bayes is a step in this direction, but we have shown that incorporating constraint-type information can lead to difficulties in interpreting objective posterior probabilities.

We believe the ultimate solution may be obtained by following what we call the fundamental principle of inferential problem solving in Section 1. In general, Fisher's fiducial argument seems to agree with this principle and applying DS theory to predictive random sets is useful for representing our uncertainty. However, because DS is subjective, one must be careful in applying DS theory when a frequency interpretation is required.

We proposed a new solution to constrained statistical inference that is consistent with the frequency interpretation. Although we chose to focus on a relatively simple example, at least, technically, our approach is very promising for the general problem of constrained inference, which appears to be difficult with existing methods. The same approach can be extended to inference about θ in the *Binomial*(θ) model, which Karl Pearson (1920) called the fundamental problem of practical statistics. We refer to Dempster (1966), Brown, Cai, and DasGupta (2001), and Zhang and Liu (2009) for different approaches to inference about the binomial model. The presence of constraints on θ will certainly make the problem more interesting and challenging.

A related problem concerns prediction of the next observation, denoted by Y, from $N(\theta,1)$ based on the observation X, where θ is unknown. In general, applying the Maximal Belief approach amounts to considering two independent realizations,

$$Z_1^* = X - \theta^* \quad \text{and} \quad Z_2^* = Y - \theta^*,$$

from the the standard normal distribution, $N(0,1)$, with complete ignorance about θ^*. A simple method is to consider $Y - X \sim N(0,2)$, which effectively integrates out θ^*. That is, $Y|X \sim N(X,2)$. A DSM can be specified for inference about Y given X in the same way as inference about θ^* based on $\theta|X \sim N(X,1)$. However, this approach may not be efficient when θ has a constraint. For example, if $\theta \in \{a,b\}$ with $b \gg a$, then observing X is approximately equivalent to observing θ. As a result, inference about Y can be based on $Y|X \sim N(a,1)$ or $N(b,1)$, depending on whether the observed value of X is close to a or b.

ACKNOWLEDGMENTS

The authors thank the chief editor, Dr. Munir Ahmad, for inviting us to contribute to this Silver Jubilee volume of PJS. We also thank the anonymous referee for the constructive comments and helpful suggestions. Jun Hui's work is sponsored by the China Scholarship Council (CSC) and the Science Foundation of the Ministry of Education of China (no. 309017).

REFERENCES

1. Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.*, 16, 101–133. with comments and a rejoinder by the authors.

2. Dawid, A. P. and Stone, M. (1982). The functional-model basis of fiducial inference, *Ann. Statist.*, 10, 1054–1074. with discussions by G. A. Barnard and by D. A. S. Fraser, and a reply by the authors.

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

4. Dempster, A. P. (2008). Dempster-Shafer calculus for statisticians, *Internat. J. of Approx. Reason.*, 48, 265–377.

5. Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics, *Philos. Trans. Roy. Soc. London Ser. A.*, 222, 309–368.

6. Fisher, R. A. (1959). *Statistical methods and scientific inference* (2nd ed.). Hafner Publishing Co. New York.

7. Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem, *J. Roy. Statist. Soc. Ser. B.*, 20, 102–107.

8. Martin, R. G., Zhang, J., and Liu, C. (2009). Situation-Specific Inference using the Dempster-Shafer Theory, Preprint. http://www.stat.purdue.edu/research/fsi/pubs/.

9. Zhang, J. and Liu, C. (2009). Dempster-Shafer inference with weak beliefs, Preprint. http://www.stat.purdue.edu/research/fsi/pubs/.

10. Pearson, K. (1920). The fundamental problem of practical statistics, *Biometrika*, 13, 1–16.

11. Zabell, S. L. (1992). R. A. Fisher and the fiducial argument, *Statist. Sci.*, 7, 369–387.