Statistica Sinica **21** (2011), 475-494

DEMPSTER-SHAFER INFERENCE WITH WEAK BELIEFS

Jianchun Zhang and Chuanhai Liu

Purdue University

Abstract: The work of A. P. Dempster in 1960s extending Fisher's fiducial inference for parametric inference using multivalued mapping, and that of G. Shafer in 1970s on the assessment and combination of evidence led to what is now known as the Dempster-Shafer (DS) theory of belief functions. However, application of DS for parametric inference has been limited due, perhaps, to its computational difficulty, non-uniqueness, and lack of frequency properties. In this paper, we return to Dempster's original approach to constructing belief functions for parametric inference, called basic DS models (BDSMs), which are usual probability models on the space of the so-called focal elements. We propose to modify BDSMs by enlarging focal elements to obtain belief functions that have desired frequency properties. We call our method Weak Belief (WB). When it enlarges the focal elements no more than necessary, the method of WB is called Maximal Belief (MB). The MB method is illustrated with two examples: (*i*) inference about a binomial proportion, and (*ii*) inference about the number of outliers ($\mu_i \neq 0$) based on the observed data $X_1, \ldots, X_n$ with the model $X_i \overset{ind}{\sim} N(\mu_i, 1)$.

Key words and phrases: Belief functions, fiducial inference, frequentist evaluation, hypothesis testing, maximal belief, predictive random sets.

1. Introduction

Dempster (1966) extended Fisher's fiducial argument to cases with multinomial observable variables and launched what we now call the DS theory of belief functions. Dempster (1967a,b, 1968a,b, 1969) applied DS to a class of statistical models, but he dropped this line of work because it could not be implemented computationally at the time. Shafer (1973, 1976) took up the theory starting in the 1970s, emphasizing the assessment and combination of evidence in general, rather than statistical modeling. DS migrated from Shafer's work to artificial intelligence via the expert systems of the time, and thence to a variety of engineering applications. Many of Dempster's and Shafer's articles, along with other classic DS articles, were recently reprinted in Yager and Liu (2008).

In the 1980s, the methodology advanced with the recognition that DS models, as well as other probabilistic and fuzzy models, could be adapted to join trees (Shenoy and Shafer (1986); Dempster (1990); Almond (1995)), where computations of marginal inferences can be reduced to local computations in a small number of dimensions. In this context, a wide variety of hidden Markov models and other network models can be regarded as DS models. Other DS models continue to be used, along with fuzzy methods, in a wide variety of engineering problems.

However, DS has not yet been accepted in the statistical community for statistical inference from observed data. This, perhaps, is due to its computational difficulty, non-uniqueness, and lack of frequency properties. For example, the application of the multinomial DS model (Dempster (1966)) and the Poisson DS model (Dempster (2008)), both proposed as general tools to build belief functions for parametric inference, has proved mathematically and computationally difficult (see, e.g., Denoeux (2006)). But recent advances in Markov chain Monte Carlo methods for Bayesian computation make DS computation possible.

For prediction of future observations from the multinomial model, Denoeux (2006) suggests a different way of building belief functions that have certain frequency properties. He proposed to build belief functions based on frequentist simultaneous confidence intervals for multinomial proportions. This idea is useful and can be viewed as an example of the general method of building belief functions based on likelihood functions. For an alternative, here we consider building belief functions by reasoning from the assumptions made in postulated sampling models.

We modify Dempster's original approach to obtain posterior belief functions that have desired frequency properties. Given a statistical model with a parameter space Θ and observation space $\mathcal{X}$, Dempster's original approach is to set up a multivalued mapping M from a probability space $\mathcal{U}$ into the product space $\Theta \times \mathcal{X}$, called the state space model (SSM). We derive the multivalued mapping from a mapping a from $\Theta \times \mathcal{U}$ to $\mathcal{X}$:

$$X = a(\theta, U) \qquad (X \in \mathcal{X}, \theta \in \Theta, U \sim \mathrm{U}(\mathcal{U})), \tag{1.1}$$

where $\mathrm{U}(\mathcal{U})$ denotes the uniform distribution in the n-dimensional cube $\mathcal{U} = [0,1]^n$, but can be replaced with any fixed distribution for generality. We call the variable U the *auxiliary variable* and (1.1) the *auxiliary (a)-equation*. The subsets

$$M(U) = \{(\theta, X) : \theta \in \Theta, X \in \mathcal{X}, X = a(\theta, U)\}, \quad U \in \mathcal{U},$$

are known as focal elements in the DS theory. Thus the probability model $U \sim \mathrm{U}(\mathcal{U})$ and the multivalued mapping $M(U)$ define a DS model (DSM) on $\Theta \times \mathcal{X}$:

$$M(U) = \{(\theta, X) : \theta \in \Theta, X \in \mathcal{X}, X = a(\theta, U)\} \qquad (U \sim \mathrm{U}(\mathcal{U})). \tag{1.2}$$

In general, a DSM on a space Ω is a usual probability model on 2^Ω, the power space of Ω consisting of all subsets of Ω. Thus, $M(U)$ is referred to as a random set when $U \sim \mathrm{U}(\mathcal{U})$ and $M(U) \neq \emptyset$.

The setting (1.1) is similar to Fisher's fiducial argument, e.g., in the context of the functional models of Bunke (1975) and Dawid and Stone (1982), and the structural inference of Fraser (1966). To call attention to the difference between (1.1) and the setting for fiducial inference, we note that (*i*) $X = a(\theta, U)$ determines a multivalued mapping from $\mathcal{U}$ to $\Theta \times \mathcal{X}$, and (*ii*) X can be the whole sample of data, rather than a (minimal) sufficient statistic as required by Fisher's fiducial argument.

For statistical inference, the a-equation (1.1) is specified in such a way that it would reproduce the probability distribution for the observed data $X \in \mathcal{X}$ when restricted to $\theta \in \Theta$. When conditioned on X, the DSM (1.2) defines the random set

$$M_X(U) = \{\theta : \ X = a(\theta, U), \, \theta \in \Theta\} \qquad (U \sim \mathrm{U}(\mathcal{U})) \tag{1.3}$$

and, thereby, a DSM on Θ for inference about θ. We call the DSM (1.3) the posterior DSM (PDSM). In the case that $M_X(U)$ is not a singleton, we "don't know" the exact value of θ in $M_X(U)$. We note that Hannig (2006) discussed the use of multivalued mappings in the context of generalized fiducial intervals, where "don't know" is removed by taking θ to be a single point in $M_X(U)$.

Let $\mathcal{A} \subseteq \Theta$ represent an assertion of interest about θ and let $\overline{\mathcal{A}}$ denote the denial of $\mathcal{A}$, *i.e.*, $\overline{\mathcal{A}} = \Theta \setminus \mathcal{A}$. Write

$$p_X(\mathcal{A}) = \frac{\Pr\left(M_X(U) \subseteq \mathcal{A}\right)}{\Pr\left(M_X(U) \neq \emptyset\right)}, \quad q_X(\mathcal{A}) = \frac{\Pr\left(M_X(U) \subseteq \overline{\mathcal{A}}\right)}{\Pr\left(M_X(U) \neq \emptyset\right)}, \tag{1.4}$$

and $r_X(\mathcal{A}) = 1 - p_X(\mathcal{A}) - q_X(\mathcal{A})$. Then, using the new terms introduced by Dempster (2008) for statisticians, we call $p = p_X(\mathcal{A})$ the probability for the truth of $\mathcal{A}$, $q = q_X(\mathcal{A})$ the probability against the truth of $\mathcal{A}$, and $r = r_X(\mathcal{A})$ the probability of "don't know", which supports neither $\mathcal{A}$ nor $\overline{\mathcal{A}}$. For readers who are familiar with Shafer (1976), we note that p is the lower probability or belief for the truth of $\mathcal{A}$ and $p + r$ is the upper probability or plausibility for the truth of $\mathcal{A}$. In the remainder of this paper, we refer to DSMs as belief models or simply beliefs.

DS (p, q, r) probabilities are personal and may not have desired frequency properties. We call the DSM (1.2) a Basic DSM (BDSM). To obtain the desired frequency property, we propose to modify the BDSM by enlarging its focal elements before conditioning on the observed data X. We do this enlargement in a systematic way, and enlarge just enough to obtain the desired frequency property. Because enlarged focal elements result in DSMs representing weaker beliefs, we call our method Weak Belief (WB). Accordingly, the WB method enlarging focal elements no more than necessary is called Maximal Belief (MB).

The remainder of this article is arranged as follows. Section 2 gives a brief introduction to the ideas from DS theory. Section 3 describes the WB and MB methods. Section 4 presents a specific class of WB models. Sections 5 and 6 illustrate the method of MB with the binomial and the many-normal-means problems. Section 7 concludes with a brief discussion.

2. A Brief Introduction to BDSMs

We review in Section 2.1 the DS calculus (Dempster (2008)) for deriving the sampling model, called sampling DSM, for data X given parameter θ and the posterior DSM for θ conditional on X, and give in Section 2.2 two illustrative examples. We assume basic knowledge of the DS calculus; See Shafer (1976),

2.1. Sampling and posterior DSMs

The sampling distribution of X given θ can be recovered by combining a DSM on $\Theta\times\mathcal{X}$ with a DSM that has the single focal element $\{(\theta, X) : X \in \mathcal{X}\} \subseteq \Theta\times\mathcal{X}$. The random set of the combined DSM is obtained, by applying Dempster's rule of combination, as the intersection of the subset $\{(\theta, X) : X \in \mathcal{X}\}$ and the random set $M(U)$ of the BDSM (1.2). It can be written as

$$\{(\theta, X) : X \in \mathcal{X}, X = a(\theta, U)\} \qquad (U \sim \mathrm{U}(\mathcal{U})).$$

Applying the DS marginalization operation on this combined DSM leads to the DSM on $\mathcal{X}$, called the sampling DSM, having the random set

$$M_\theta(U) = \{X : X \in \mathcal{X}, X = a(\theta, U)\} \qquad (U \sim \mathrm{U}(\mathcal{U})). \tag{2.1}$$

Similarly, one can derive the posterior DSM (1.3) discussed in Section 1.

2.2. Examples

Example 2.1. In this example, we consider the simple Gaussian model with the a-equation

$$X = \mu + \Phi^{-1}(U) \qquad (\mu \in \mathcal{R},\, U \sim \mathrm{U}(0, 1)), \tag{2.2}$$

where $\mathcal{R} = (-\infty, \infty)$ and $\Phi^{-1}(\cdot)$ stands for the inverse CDF of the standard normal distribution $N(0, 1)$. That is, the sampling model is $X \sim N(\mu, 1)$ with unknown $\mu \in \mathcal{R}$. The SSM is the product space $\mathcal{R} \times \mathcal{R}$ for (μ, X). The focal elements are the lines $M(U) = \{(\mu, X) :\ X = \mu + \Phi^{-1}(U)\}$ indexed by $U \in [0, 1]$. Routine application of DS calculus leads to the following results: (i) the sampling distribution of X given μ is $N(\mu, 1)$, and (ii) the posterior DSM for μ given X is the usual fiducial distribution $\mu|X \sim N(X, 1)$.

Example 2.2. Let X be a dichotomous observation with $X \in \mathcal{X} = \{0, 1\}$. Suppose that the Bernoulli model Bernoulli(θ)

$$\Pr_\theta(X = 1) = \theta \qquad \text{and} \qquad \Pr_\theta(X = 0) = 1 - \theta \tag{2.3}$$

with unknown $\theta \in \Theta = [0,1]$ is considered to generate the observed data X. The problem is to infer θ from X. We use the following a-equation for the quantities X and θ with an auxiliary random variable $U \sim \mathrm{U}(0,1)$:

$$X = \begin{cases} 1, & \text{if } U \leq \theta; \\ 0, & \text{if } U > \theta. \end{cases} \tag{2.4}$$

It follows that the sampling DSM has the random set

$$M_\theta(U) = \begin{cases} \{1\}, & \text{if } U \leq \theta; \\ \{0\}, & \text{if } U > \theta, \end{cases} \qquad (U \sim \mathrm{U}(0,1)) \tag{2.5}$$

which is consistent with the Bernoulli model (2.3). The posterior DSM for θ has the random set

$$M_X(U) = [U,1] \text{ for } X = 1, \text{ and } [0,U] \text{ for } X = 0 \quad (U \sim \mathrm{U}(0,1)). \tag{2.6}$$

To illustrate the DS (p,q,r) output based on the DSM (2.6) with state space $\Theta = [0,1]$, consider the assertion $\mathcal{A} = \{\theta \leq \theta_0\} \subseteq \Theta$ for a known θ_0. Given $X = 1$, for example, we have the random interval $[U,1]$ for θ with $U \sim U(0,1)$. There are two possible cases: (i) the case of $U > \theta_0$, which provides evidence against the truth of $\mathcal{A}$, and (ii) the case of $U \leq \theta_0$, which does not have any information about the truth or falsity of $\mathcal{A}$. Note that there are no realizations of the random interval that provide evidence for the truth of $\mathcal{A}$. As a result, the DS output for the assertion $\mathcal{A}$ has the following (p,q,r) components

$$p_X(\mathcal{A}) = 0, \quad q_X(\mathcal{A}) = \Pr(U > \theta_0) = 1 - \theta_0, \text{ and } r_X(\mathcal{A}) = \theta_0.$$

3. Weak and Maximal Beliefs

Suppose that the a-equation (1.1) is considered for making inference about an unknown θ given the observed data X. We are interested in making inference about an assertion $\mathcal{A} \subseteq \Theta$.

3.1. Credibility: a frequentist evaluation

DS inference would be questionable if large values of $p_X(\mathcal{A})$ under the truth of $\overline{\mathcal{A}}$ or large values of $q_X(\mathcal{A})$ under the truth of $\mathcal{A}$ occur frequently in repeated experiments. This motivates the following definition of credibility of DS inference.

Definition 3.1. Suppose that the observed data model X is specified by the a-equation (1.1) with unknown $\theta \in \Theta$. Given $\alpha \in (0,1)$, the DS $(p_X(\mathcal{A}), q_X(\mathcal{A}), r_X(\mathcal{A}))$ output for an assertion $\mathcal{A}$ is said to be credible at α-level if

$$\Pr_\theta(p_X(\mathcal{A}) \geq 1 - \alpha) \leq \alpha \tag{3.1}$$

for every $\theta \in \overline{\mathcal{A}}$ and

$$\Pr_\theta\left(q_X(\mathcal{A}) \geq 1-\alpha\right) \leq \alpha \tag{3.2}$$

for every $\theta \in \mathcal{A}$, where the distribution of the random variable X is determined by the a-equation (1.1) and $\theta \in \Theta$. The DS $(p_X(\mathcal{A}), q_X(\mathcal{A}), r_X(\mathcal{A}))$ output for an assertion $\mathcal{A}$ is said to be credible if (3.1) and (3.2) hold for all $\alpha \in (0,1)$.

To explain the definition of credibility, we consider the following simple solution to the problem of choosing $\mathcal{A}$, $\overline{\mathcal{A}}$, or neither, given the observed data X. Take a small value α, e.g., $\alpha = 0.05$, and choose $\mathcal{A}$ if $p > 1-\alpha$, $\overline{\mathcal{A}}$ if $q > 1-\alpha$, and neither $\mathcal{A}$ nor $\overline{\mathcal{A}}$ otherwise. It follows that if the (p,q,r) is credible at α-level, we would make wrong choices at most $\alpha \times 100\%$ of the times. A familiar such example is the precise/sharp hypothesis problem. Incidentally, we note that the above definition of credibility is related to the fundamental frequentist principle of Walley (2002).

Example 3.1. Consider the Gaussian model $N(\mu,1)$ with unknown mean μ in Example 2.1 for a single observation X. Example 2.1 showed that the PDSM for inference about μ is the familiar fiducial posterior, *i.e.*, $\mu|X \sim N(X,1)$. Here we consider the assertions

(i) $\mathcal{A}_1 = \{\mu \leq \mu_0\}$ for fixed μ_0, and
(ii) $\mathcal{A}_2 = \{\mu_0 - \delta \leq \mu \leq \mu_0 + \delta\}$ for fixed μ_0 and $\delta \geq 0$.

The DS (p,q,r) output for the assertion $\mathcal{A}_1$ is given by

$$p_X(\mathcal{A}_1) = \Pr\left(\mu \leq \mu_0 | X\right) = \Phi(\mu_0 - X), \quad q_X(\mathcal{A}_1) = 1 - p_X(\mathcal{A}_1),$$

and $r_X(\mathcal{A}_1) = 0$. For any $\alpha \in (0,1)$, we have for every $\mu \in \overline{\mathcal{A}_1}$, *i.e.*, $\mu > \mu_0$,

$$\begin{aligned}\Pr_\mu\left(p_X(\mathcal{A}_1) \geq 1-\alpha\right) &= \Pr_\mu\left(\Phi(\mu_0 - X) \geq 1-\alpha\right)\\ &= \Pr_\mu\left(X \leq \mu_0 - \Phi^{-1}(1-\alpha)\right)\\ &= \Phi(\mu_0 - \mu - \Phi^{-1}(1-\alpha))\\ &\leq \Phi(\Phi^{-1}(\alpha)) = \alpha\end{aligned}$$

and similarly for every $\mu \in \mathcal{A}_1$, $\Pr_\mu\left(q_X(\mathcal{A}_1) \geq 1-\alpha\right) \leq \alpha$. Thus, the fiducial inference about $\mathcal{A}_1$ is credible for all $\alpha \in (0,1)$.

For $\mathcal{A}_2$, the DS (p,q,r) output has the components

$$p_X(\mathcal{A}_2) = \Phi(\mu_0 + \delta - X) - \Phi(\mu_0 - \delta - X), \quad q_X(\mathcal{A}_2) = 1 - p_X(\mathcal{A}_2),$$

and $r_X(\mathcal{A}_2) = 0$. It follows that for $\delta \approx 0$ and $\alpha \in (0,1)$,

$$\Pr_\mu\left(q_X(\mathcal{A}_2) \geq 1-\alpha\right) = \Pr_\mu\left(\Phi(\mu_0 + \delta - X) - \Phi(\mu_0 - \delta - X) \leq \alpha\right) \approx 1$$

for every $\mu \in \mathcal{A}_2$. This result shows that the DS inference about the assertion $\mathcal{A}_2$ with a small δ is not credible.

3.2. Weakening BDSMs: a motivating example

The posterior DSM for inference about the unknown parameter μ in the Gaussian model $N(\mu, 1)$ from a single observation X may fail to be credible. This indicates that the belief specified in the BDSM to derive the posterior DSM is too strong in the sense that the resulting p or q are too large in repeated experiments for certain assertions. To take a closer look at what that belief is, assume that the observation X was indeed generated according to (2.2). In this case, inference about the unknown θ is the same as inference about the unobserved realization of U in the specific experiment. Let U^* denote this unobserved realization of U. Then, U^* is known to have followed $\mathrm{U}(0,1)$ and satisfies the identity

$$X = \mu + \Phi^{-1}(U^*). \tag{3.3}$$

The fiducial distribution $\mu|X \sim N(X, 1)$ can be viewed as obtained from (3.3) by predicting U^* with a random draw U from $\mathrm{U}(0,1)$. We call the random variable U the *predictive random variable* (PRV) and we call U^* the *generative random variable* (GRV). The BDSM for posterior inference is effectively specified by assigning the distribution of the GRV to the PRV.

For credible DS inference with the Gaussian model, we weaken the BDSM and, thereby, the posterior DSM by expanding U into an interval, denoted by $S(U)$. To illustrate the idea, we enlarge the PRV U into the random interval

$$S(U) = \left[U - \frac{U}{2},\, U + \frac{1-U}{2}\right] \qquad (U \sim \mathrm{U}(0,1)). \tag{3.4}$$

This modification replaces $\mu|X \sim N(X, 1)$ with the DSM

$$S_X(U) = \left\{\mu : X - \Phi^{-1}\left(\frac{U+1}{2}\right) \le \mu \le X - \Phi^{-1}\left(\frac{U}{2}\right)\right\} \quad (U \sim \mathrm{U}(0,1)).$$

To investigate the credibility of this modified DSM for inference about the sharp assertion $\{\mu = \mu_0\}$, for which the BDSM is not credible, we now have the DS (p, q, r) output

$$p_X(\{\mu = \mu_0\}) = 0, \quad q_X(\{\mu = \mu_0\}) = 2\Phi(|X - \mu_0|) - 1,$$

and $r_X(\{\mu = \mu_0\}) = 1 - q_X(\{\mu = \mu_0\})$. Thus, the long-run frequency distribution of $q_X(\{\mu = \mu_0\})$ is the uniform on the interval $[0,1]$ when $X \sim N(\mu_0, 1)$. It follows that for all $\alpha \in [0,1]$ $\Pr_\mu(p_X(\{\mu = \mu_0\}) \ge 1 - \alpha) = 0 \ \ (\le \ \alpha)$ for $\mu \ne \mu_0$, and $\Pr_\mu(q_X(\{\mu = \mu_0\}) \ge 1 - \alpha) = \alpha \ \ (\le \ \alpha)$ for $\mu = \mu_0$. Hence, the resulting weak belief model specified by (3.4) leads to a modified posterior DSM that is credible for the assertion $\{\mu = \mu_0\}$.

A formal definition of weak belief is given in Section 3.3. The particular choice of the above random interval is related to the concept of maximal belief of Section 3.4, and is discussed further in Section 4.

3.3. Weak beliefs

For a given DSM B and an assertion $\mathcal{A}$, a subset of the SSM for B, we write the components of (p, q, r) for $\mathcal{A}$ as $(p_B(\mathcal{A}), q_B(\mathcal{A}), r_B(\mathcal{A}))$. This notation is consistent with (1.4) in the sense that the observed data X in (1.4) indexes different DSMs. Let $S \sim B$, that is, S is the random set of B. Then

$$p_B(\mathcal{A}) = \Pr(S \subseteq \mathcal{A}), \quad q_B(\mathcal{A}) = \Pr\left(S \subseteq \overline{\mathcal{A}}\right), \tag{3.5}$$

and $r_B(\mathcal{A}) = 1 - p_B(\mathcal{A}) - q_B(\mathcal{A})$. One more useful DS concept is the so-called commonality function:

$$c_B(\mathcal{A}) = \Pr(S \supseteq \mathcal{A}), \tag{3.6}$$

which was introduced by Shafer (1976) and plays an important role in DS calculus.

For building credible DSMs, we consider beliefs that are weaker than the BDSM.

Definition 3.2. Let B and B' be two DSMs on a common SSM. The DSM B is said to be *weaker* than the DSM B' if $p_B(\mathcal{A}) \leq p_{B'}(\mathcal{A})$ holds for every assertion $\mathcal{A}$.

For convenience, a belief is said to be *weak* if it is weaker than the corresponding BDSM. Weak beliefs can be interpreted from different perspectives that are summarized in the following three propositions, where all DSMs are assumed to be on a common SSM. Proposition 3.1 serves as an alternative definition in terms of commonality. Proposition 3.2 implies that weaker DSMs have a larger probability of "don't know". Proposition 3.3 provides a sufficient condition for comparing the weakness of two beliefs and suggests a way of creating weaker beliefs. The proofs of these results are straightforward and therefore omitted here to save space.

Proposition 3.1. *Suppose that B and B' are two DSMs on a common SSM. If $c_B(\mathcal{A}) \geq c_{B'}(\mathcal{A})$ holds for every assertion $\mathcal{A}$, then B is weaker than B'.*

Proposition 3.2. *If the DSM B is weaker than the DSM B', then $r_B(\mathcal{A}) \geq r_{B'}(\mathcal{A})$ for every assertion $\mathcal{A}$.*

Proposition 3.3. *Let S and S' be the random sets of the DSMs B and B', respectively. If the random set S can be obtained via a mapping $S = m(S')$ in such a way that $S' \subseteq S = m(S')$, then B is weaker than B'.*

Let U^* be the realization of U that corresponds to the observed data X via the a-equation $X = a(\theta, U^*)$. Let B_0 denote the BDSM for predicting U^*. To weaken B_0, we make use of Proposition 3.3 and replace the PRV U with a subset $S(U)$ of $\mathcal{U}$ containing U, *i.e.*, $U \in S(U)$. Accordingly, the posterior DSM (1.3) becomes the weak DSM that has the random set

$$M_{X,S}(U) = \{\theta : \theta \in \Theta, X = a(\theta, u) \text{ for some } u \in S(U)\}, \tag{3.7}$$

where $U \sim \mathrm{U}(\mathcal{U})$. Thus, the (p, q, r) output produced by the weakened DSM for any assertion $\mathcal{A} \subseteq \Theta$ has the p, q, r-components

$$p_{X,S}(\mathcal{A}) = \frac{\Pr\left(M_{X,S}(U) \subseteq \mathcal{A}\right)}{\Pr\left(M_{X,S}(U) \neq \emptyset\right)}, \quad q_{X,S}(\mathcal{A}) = \frac{\Pr\left(M_{X,S}(U) \subseteq \overline{\mathcal{A}}\right)}{\Pr\left(M_{X,S}(U) \neq \emptyset\right)}, \tag{3.8}$$

and $r_{X,S}(\mathcal{A}) = 1 - p_{X,S}(\mathcal{A}) - q_{X,S}(\mathcal{A})$.

3.4. The method of maximal belief

Weak beliefs introduced in Section 3.3 are not unique. Assuming a class of such weak beliefs of interest is available, we can seek a particular belief within the class to balance between credibility and efficiency.

Let U^* be an unobserved realization of $U \sim \mathrm{U}(\mathcal{U})$ and let B be a DSM with the random set S, called the predictive random set (PRS), for inference about U^*. Let

$$m_B(U^*) = \Pr\left(S \not\ni U^*\right). \tag{3.9}$$

For credible inference, we want to bound the frequency of large values of $m_B(U^*)$. This motivates the following definition of credibility of beliefs for predicting U^*.

Definition 3.3. Given $\alpha \in (0, 1)$, a belief B for inferring (or predicting) U^* is said to be credible at level α if

$$\Pr\left(m_B(U^*) \geq 1 - \alpha\right) \leq \alpha, \tag{3.10}$$

where $U^* \sim \mathrm{U}(\mathcal{U})$. A belief B for inferring (or predicting) U^* is said to be credible if it is credible at level α for all $\alpha \in [0, 1]$.

The following result relates the credibility of a PRS $S(U)$, where $U \sim \mathrm{U}(\mathcal{U})$, for predicting U^* and the credibility of the corresponding DS (p, q, r) output for assertions about θ.

Theorem 3.1. *Suppose that the BDSM is defined by the focal elements* (1.2) *with* $U \sim U(\mathcal{U})$*. If a random set* $S(U)$ *with* $U \sim U(\mathcal{U})$ *is credible at* α*-level for predicting* U^**, a realization from* $U(\mathcal{U})$*, and* $Pr(M_X(U) = \emptyset) = 0$*, then the DS* (p, q, r) *output* (3.8) *for every assertion* $\mathcal{A} \subseteq \Theta$ *is credible at* α*-level.*

Proof. Let $\mathcal{A}$ be any assertion of interest. Then the probability $q_{X,S}(\mathcal{A})$ against the truth of $\mathcal{A}$ is smaller than the probability $q_{X,S}(\{\theta\})$ for every $\theta \in \mathcal{A}$, which follows from

$$q_{X,S}(\mathcal{A}) = \Pr\left(M_{X,S}(U) \subseteq \Theta \setminus \mathcal{A}\right) \leq \Pr\left(M_{X,S}(U) \subseteq \Theta \setminus \{\theta\}\right) = q_{X,S}(\{\theta\}), \tag{3.11}$$

where the two equalities in (3.11) follow the assumption $\Pr(M_X(U) = \emptyset) = 0$. Note that the event $M_{X,S}(U) \subseteq \Theta \setminus \{\theta\}$ is equivalent to $\theta \notin M_{X,S}(U)$, that is, there is no $u \in S(U)$ such that $a(\theta, u) = X$. This implies that $U^* \notin S(U)$ because U^* is known to satisfy $a(\theta, U^*) = X$. Thus, it follows from (3.11) that

$$q_{X,S}(\mathcal{A}) \leq q_{X,S}(\{\theta\}) \leq \Pr(S(U) \not\ni U^*).$$

That is, $q_{X,S}(\mathcal{A})$ is stochastically smaller than $\Pr(S(U) \not\ni U^*)$ in repeated experiments. Making use of the condition that $S(U)$ is credible for predicting U^* at α-level, (3.2) holds for $q_{X,S}(\{\theta\})$. The symmetry argument based on $p_{X,S}(\mathcal{A}) = q_{X,S}(\overline{\mathcal{A}})$ with $\theta \in \overline{\mathcal{A}}$ leads to the conclusion that (3.1) holds for $p_{X,S}(\{\theta\})$. This completes the proof.

Among all beliefs credible at level α, some can be more efficient than others. In general, the smaller the coverage probability $\Pr(S \ni U^*)$, the more efficient the belief B with the PRS S. Note that $\Pr(S \ni U^*) = 1 - m_B(U^*)$; See (3.9). This motivates the definition of a *maximal belief* (MB) at level α with respect to a class of beliefs.

Definition 3.4. Let B_α be a class of beliefs that are credible at level α. A belief $B \in \mathsf{B}_\alpha$ is said to be a maximal belief at level α with respect to the class B_α if

$$\Pr(m_B(U^*) \geq 1 - \alpha) = \max_{B' \in \mathsf{B}_\alpha} \Pr\left(m_{B'}(U^*) \geq 1 - \alpha\right). \tag{3.12}$$

The following results are useful for constructing MBs.

Proposition 3.4. *If a belief* B *on the SSM* $\mathcal{U}$ *satisfies*

$$Pr(m_B(U^*) \geq 1 - \alpha) = \alpha \quad (U^* \sim U(\mathcal{U})), \tag{3.13}$$

then it is an MB. Furthermore, if (3.13) *holds for all* $\alpha \in (0, 1)$*, then* $m_B(U^*) \sim U(0, 1)$*.*

The discussion in previous sections is on the credibility and efficiency of DSMs and their weakened versions for all assertions. For a given assertion $\mathcal{A}$ of interest, we can find an assertion-specific belief B such that it is both credible and efficient for inference about $\mathcal{A}$. For example, for the assertion $\{\mu \leq \mu_0\}$ in the Gaussian model $X \sim N(\mu, 1)$ with the observed data X, the BDSM is both credible and efficient. Section 6 provides another example of using assertion-specific WBs.

4. A Class of Predictive DSMs for Uniform Samples

In this section we present a particular class of PRSs, based on intuition and geometric simplicity, for predicting an unobserved sample from the uniform distribution $\mathrm{U}(0,1)$. The corresponding class of weak beliefs is used in Sections 5 and 6 to illustrate the proposed MB method.

4.1. A class of predictive DSMs for a single uniform random variable

For each point U in $[0,1]$, we consider the subset of the form

$$S_w(U) = [U - wU,\ U + w(1-U)] \qquad (w \in [0,1]).$$

Let $U \sim \mathrm{U}(0,1)$. Then we have a class of beliefs indexed by $w \in [0,1]$:

$$\mathsf{B} = \{B_w : 0 \leq w \leq 1\},$$

where the belief B_w has the random set $S_w(U)$. Note that the interval length of $S_w(U)$ is w. Thus, $B_0(U)$ is the BDSM used for fiducial inference while $B_1(U)$ represents the vacuous belief that has the entire space as the single focal element. It can be shown that the MB for any level α is $B_{1/2}$, which has the random set

$$S_{1/2}(U) = \left[\frac{U}{2}, \frac{U+1}{2}\right] \qquad (U \sim \mathrm{U}(0,1)).$$

Example 4.1. Consider again the Gaussian example of Section 3.2 with a single observation X from $N(\mu, 1)$ with unknown mean μ. Here we conclude this "running" example with some numerical results. The random interval of the MB for μ can be written as

$$M_{X,1/2}(U) = [X - \Phi^{-1}(U + \frac{1}{2}), X - \Phi^{-1}(U)] \qquad (U \sim U(0, \frac{1}{2})).$$

For the assertion $\mathcal{A} = \{\mu = \mu_0\}$ with fixed $\mu_0 \in \mathcal{R}$, the probability p for the truth of $\mathcal{A}$ is 0 due to the fact that $\Pr\left(X - \Phi^{-1}(U + 1/2) = X - \Phi^{-1}(U) = \mu_0\right) = 0$. The probability q against the truth of $\mathcal{A}$ is $q = 2\Phi(|X - \mu_0|) - 1$, the probability that the random interval $M_{X,1/2}(U)$ does not contain μ_0. For example, for $\mu_0 = 0$ with the observed $X = 0$ we have $(p, q, r) = (0, 0, 1)$, which indicates no evidence for or against the truth of the assertion that $\mu = 0$. For $\mu_0 = 0$ with the observed $X = 2$ we have $(p, q, r) = (0, 0.95, 0.05)$, which shows evidence with $q = 95\%$ against the truth of the assertion that $\mu = 0$. This demonstrates a nice DS way of resolving the problem of significance testing with the null hypothesis $H_0 : \mu = 0$ and the alternative hypothesis $H_a : \mu \neq 0$. DS (p, q, r) outputs for other assertions can also be computed similarly. For example, for the assertion $\mu \leq 0$ we have $(p, q, r) = (0, 0, 1)$ conditional on the observed data $X = 0$ and $(p, q, r) \approx (0, .95, 0.05)$ conditional on $X = 2$.

4.2. A class of *predictive* DSMs for ordered uniforms

For a uniform sample $U_1, \ldots, U_n$, we write ordered values as $U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)}$. A draw of $U_{(1)}, \ldots, U_{(n)}$ from the BDSM for $U_{(1)}, \ldots, U_{(n)}$ can be obtained by taking a sample of n from $U(0, 1)$ and sorting the sample in ascending order. For large n, a more efficient method of generating $U_{(1)}, \ldots, U_{(n)}$ is to take a sample of $n + 1$, denoted by $Z_1, \ldots, Z_{n+1}$, from the standard exponential distribution Expo(1), with $U_{(i)} = \sum_{j=1}^{i} Z_j / \sum_{j=1}^{n+1} Z_j$ for $i = 1, \ldots, n$.

It is known that the marginal distribution of $U_{(i)}$ is the Beta distribution Beta$(i, n - i + 1)$ for $i = 1, \ldots, n$. To construct a random set for predicting an unobserved realization, denoted by $U^*_{(1)}, \ldots, U^*_{(n)}$, we consider replacing $U_{(i)}$ of a random draw $U_{(1)}, \ldots, U_{(n)}$, with an interval. The upper end point of the interval is set to the κ $(0 \leq \kappa \leq 1)$ quantile of the truncated distribution Beta$(i, n-i+1)$ restricted to the interval from $U_{(i)}$ to 1. The lower end point of the interval is set to the $(1 - \kappa)$ quantile of the truncated distribution Beta$(i, n - i + 1)$ restricted to the interval $[0, U_{(i)}]$.

Let $\mathcal{O}_n = \{(v_1, \ldots, v_n) : 0 \leq v_1 \leq \cdots \leq v_n \leq 1\}$. Formally, in terms of a DSM we define the focal element as $K(V, \kappa) = \{v : v \in \mathcal{O}_n$ and $A_i(V_i, \kappa) \leq v_i \leq B_i(V_i, \kappa)$ for all $i = 1, \ldots, n\}$ where $V \in \mathcal{O}_n$, $\kappa \in [0, 1]$, $A_i(V_i, \kappa) = \text{qBeta}(P_i(V_i) - \kappa P_i(V_i), i, n-i+1)$ and $B_i(V_i, \kappa) = \text{qBeta}(P_i(V_i) + \kappa(1 - P_i(V_i)), i, n-i+1)$, with $P_i(V_i) = \text{pBeta}(V_i, i, n-i+1)$ for $i = 1, \ldots, n$. The functions pBeta$(\cdot, i, n-i+1)$ and qBeta$(\cdot, i, n - i + 1)$ stand for the CDF of Beta$(i, n - i + 1)$ and the inverse CDF of Beta$(i, n - i + 1)$. We define a measure on the focal element space as follows.

1. $V = (V_1, \ldots, V_n)$ and κ are independent,
2. $V = (V_1, \ldots, V_n)$ follows the distribution of the ordered uniform $(U_{(1)}, \ldots, U_{(n)})$, and
3. $\kappa = 1/2 + L/2$ with $L \sim$ Beta$(w_n, 1)$ and $w_n \geq 0$.

This results in a class of DSMs with the random set S_{w_n} indexed by w_n.

The use of a distribution for κ is motivated by the fact that in the general n case, there does not exist a constant $\kappa \in [0, 1]$ that produces a satisfactory

DSM for balancing credibility and efficiency. The particular choice of the class of distributions for κ is *ad hoc* and based on both mathematical simplicity and flexibility for finding a satisfactory MB model. For the $n = 1$ case, we take $w_1 = 0$, which gives the MB model discussed in Section 4.1.

Given a prespecified value α, e.g., $\alpha = 0.05$, the MB is obtained by finding a solution w_n to

$$\Pr\left(m_{w_n}(U^*_{(1)}, \ldots, U^*_{(n)}) \geq 1 - \alpha\right) = \alpha. \tag{4.1}$$

For any fixed w_n, $m_{w_n}(U^*_{(1)}, \ldots, U^*_{(n)})$ can be simulated using Monte Carlo methods. Since the long-run frequency distribution of $m_{w_n}(U^*_{(1)}, \ldots, U^*_{(n)})$ is monotone in w_n, the solution w_n to (4.1) can be obtained via the Stochastic Approximation (SA) algorithm of Robbins and Monro (1951). For example, with fixed $\alpha = 0.05$, the SA algorithm-based on simulated $m_{w_n}(U^*_{(1)}, \ldots, U^*_{(n)})$ produced the following results for a set of values of n

n	1	2	3	5	10	100	1,000	10,000
w_n	0	0.33	0.57	0.98	1.8	6.6	13.7	22

It appears that w_n for n in the range from 3 to 100 is approximately linear in $(\ln n)^2$. This approximation is used in Section 6 for estimating the number of outliers in the many-normal-means problem.

4.3. A class of *predictive* DSMs for unordered uniforms

For predicting an unobserved realization $(U^*_1, \ldots, U^*_n)$ from $U([0, 1]^n)$, we make use of the random set proposed in Section 4.2 for $U^*_{(1)} \leq U^*_{(2)} \leq \cdots \leq U^*_{(n)}$, the ordered values of $U^*_1, \ldots, U^*_n$. What is needed is a permutation π ($\in \mathcal{P}_n$) that assigns $(U^*_{(1)}, ..., U^*_{(n)})$ to $(U^*_1, \ldots, U^*_n)$, $U^*_i = U^*_{(\pi_i)} \quad (i = 1, \ldots, n)$, where $\mathcal{P}_n$ is the set of the $n!$ permutations of $(1, \ldots, n)$. Mathematically, we need to specify a DSM on the space $\mathcal{P}_n$. In this paper, we consider the *vacuous* DSM that is, we take the DSM with $\mathcal{P}_n$ as the single focal element. Care must be taken, however, in computing (p, q, r) for certain assertions because there is *one and only one* unknown assignment permutation. The use of this DSM is illustrated in Section 6 for the multiple testing example.

5. The Binomial Problem

Inference about the binomial proportion θ based the observed data X from the binomial distribution Binomial(n, θ) with known size n and unknown $\theta \in [0, 1]$ is a fundamental problem of statistics (Pearson (1920); Clopper and Pearson (1934); Brown, Cai, and DasGupta (2001); and references therein). DS inference

Table 1. The (p, q, r) for the assertion $\mathcal{A} = \{\theta = 1.6\%\}$ based on the observed data X with known n in the binomial example.

Data (X, n)	(p, q, r)	Fisher's p-value
(24, 1,000)	(0, 0.9290, 0.0710)	0.0438
(1,680, 100,000)	(0, 0.9536, 0.0464)	0.0438

about θ (Dempster (1966)) provides the first classical example of DS parametric inference. When conditioned on X, the posterior DSM for θ is the random interval $[U_{(X)}, U_{(X+1)}]$, with the two end points $U_{(X)}$ and $U_{(X+1)}$ being the X-th and $(X+1)$-th order statistics of a sample of n from $\mathrm{U}(0,1)$.

Here we consider WB models based on the a-equation

$$X = a(\theta, U) \qquad (\theta \in [0,1], U \sim \mathrm{U}(0,1)), \tag{5.1}$$

where $X = a(\theta, U)$ is given the constraints

$$\sum_{k=0}^{X-1} \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k} \le U < \sum_{k=0}^{X} \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}. \tag{5.2}$$

The two bounds for U in (5.2) are the CDF values of Binomial(n, θ) evaluated at $X-1$ and X. Formally, the SSM of the DSM concerning the pair of quantities X and θ is $\{0, 1, ..., n\} \times [0, 1]$. It is easy to show that (i) the sampling DSM gives the sampling distribution Binomial(n, θ) for X given θ and n, and the posterior DSM has the random set

$$M_X(U) = \{\theta : \mathrm{qBeta}(U, X, n-X+1) \le \theta \le \mathrm{qBeta}(U, X+1, n-X)\}, \tag{5.3}$$

where $U \sim \mathrm{U}(0,1)$ and qBeta$(\cdot, \alpha, \beta)$ denotes the inverse CDF of the beta distribution Beta(α, β). It is easy to see that (5.3) is an interval and that the marginal distributions of the two end points of this random interval are the same as those of the random interval in the DSM of Dempster (1966). For WB analysis, we prefer the posterior DSM with the random set (5.3) to the DSM of Dempster (1966) because we need to predict only the univariate random variable U in (5.3) and have to predict the bivariate random variable $(U_{(X)}, U_{(X+1)})$ in the DSM of Dempster (1966).

Suppose that we use the PRS $S(U) = [U/2, (U+1)/2]$ discussed in Section 4. The WB model has the following random set

$$M_{X,S}(U) = \left\{ \begin{array}{l} \theta : \mathrm{qBeta}(u, X, n-X+1) \le \theta \le \\ \quad \mathrm{qBeta}(u, X+1, n-X) \text{ for some } u \in S(U) \end{array} \right\}, \tag{5.4}$$

where $U \sim \mathrm{U}(0,1)$. For a numerical illustration, consider the two artificial data sets (i) $n = 1,000$ and $X = 24$, and (ii) $n = 100,000$ and $X = 1,680$, which are similar to the two Poisson examples of Dempster (2008). Assume that the assertion of interest is $\mathcal{A} = \{\theta = 1.6\%\}$ in the two cases. The probability for this assertion is zero and the probability against this assertion is given by

$$\Pr\left(\mathrm{qBeta}(\frac{U}{2}, X, n-X+1) > 1.6\% \text{ or } \mathrm{qBeta}\Big(\frac{U+1}{2}, X+1, n-X\Big) < 1.6\%\right),$$

where $U \sim \mathrm{U}(0,1)$. These probabilities are shown in Table 1, where the Fisher p-values based on the normal approximation are also given. As discussed by Dempster (2008), it is interesting to see that Fisher's p-value should be interpreted as a part of r, the probability of "don't know". We note that for obtaining sensible (p, q, r) output for assertions, Dempster (2008) considered a "dull" *null*, which effectively increases the value of r. With MB, such a treatment seems to be unnecessary, making MB attractive for hypothesis testing.

6. The Many-Normal-Means Problem

We consider the many-normal-means problem $X_i \overset{ind}{\sim} N(\mu_i, 1)$ with unknown means μ_i, $i = 1, \ldots, n$. This is an important problem that we call the *second* fundamental problem of practical statistics, while referring the *first* fundamental problem to the binomial population mean problem (Pearson (1920)). Here we use it as an illustrative example by taking $n = 100$, and considering the sequence of assertions concerning the number of "outliers" ($\mu_i \neq 0$)

$$\mathcal{A}_K = \{|\{\mu_i : \ \mu_i \neq 0, i = 1, \ldots, n\}| < K\} \tag{6.1}$$

for $K = 1, 2, \ldots$, where $|S|$ denotes the number of elements in the set S.

To compute our (p, q, r) probabilities for $\mathcal{A}_K$ in (6.1), we use the a-equation

$$X_i = \mu_i + \Phi^{-1}(U_i) \qquad (U_i \overset{i.i.d.}{\sim} U(0,1),\ i = 1, \ldots, n).$$

One can use the predictive random set for $U_1, \ldots, U_n$, as discussed in Section 4. The needed technique is essentially the same as what is described below for an alternative MB method, where we are concerned with a predictive DSM for a subset of $U_1, \ldots, U_n$. The purpose here is to show that MB analysis can be conducted at the assertion level, that is, the MB analysis can be tailored for the assertion(s) of interest.

For each assertion $\mathcal{A}_K$, we have no evidence for the truth of the assertion because the posterior probability for each μ_i being zero is zero. Thus, we have $p = 0$ for all $\mathcal{A}_K$, $K = 1, 2, \ldots$ Note that the assertion $\mathcal{A}_K$ can be stated as "there are *at most* $K-1$ outliers" in $\mu_1, \ldots, \mu_n$. To compute the probability against the truth of $\mathcal{A}_K$ for each $K = 1, 2, \ldots$, we need only find evidence that there does not exist $U_{i_1}, \ldots, U_{i_{n-K}}$, a sample $n - K$ from $U(0, 1)$, such that

$$X_{i_j} = \Phi^{-1}(U_{i_j}) \qquad (j = 1, \ldots, n - K).$$

Computationally, one way of doing this is to first generate a predictive random set for the ordered $(n - K)$, instead of n, uniforms and then to assign each of the $n - K$ intervals, denoted by $[a_j, b_j]$, for $U_{i_1}, \ldots, U_{i_{n-K}}$ to at most one of the observed data $\{X_i\}_{i=1}^n$ in such a way that the number of matched interval-data assignments is *maximized*. The required maximization is due to the fact that we use the *vacuous* DSM for the unknown assignment permutation discussed in Section 4.3. The cases with unmatched intervals provide evidence against the truth of $\mathcal{A}_K$. This matching problem is a simple version of the *maximum assignment problem*. It can be solved in a straightforward manner by assigning $[a_j, b_j]$ to the smallest X_i values that satisfy $a_j \leq \Phi(X_i) \leq b_j$ in the order $j = 1, \ldots, n - K$. This method creates *greedy* matching and is known as Glover's algorithm (Glover (1967); Soares and Stefanes (2007)).

To see performance, we conducted a simulation study. To create the observed data, four types of μ_is were considered:

(a) $\mu_i = 0$ for all $i = 1, \ldots, 100$;

(b) 90 of μ_i are zero and the other 10 were generated from $2 + \text{Expo}(1)$;

(c) 90 of μ_i are zero and the other 10 were generated from $4 + \text{Expo}(1)$; and

(d) 90 of μ_i are zero and the other 10 were generated from $6 + \text{Expo}(1)$.

Each case was replicated 10 times, resulting in 10 sequences of probabilities for the truth of the assertion that *there are at least* K *outliers* for $K = 1, 2, \ldots$. These probabilities are shown in Figure 1(a)$-$(d). The fact that the probabilities for the assertion that *there is at least one outlier* are spread quite evenly along the vertical axis in Figure 1(a) shows that the MB posterior probability is approximately frequency-calibrated, which is supported by Figure 2, the histogram of the MB posterior probability obtained from a separate simulation study with 1,000 replicates of case (a). This can also be seen to some extent in Figure 1(d). Case (b) is relatively difficult for detecting outliers because intuitively, observed values in the interval, say, from 1 to 2, would cause problems. Even in this difficult case, using both large probability values and their sequential changes/differences would result in a good estimate of the number of outliers, considering that the probabilities are intended to be used only for a kind of lower bound on the number of outliers. Case (b) contains an interesting simulated data set, where all the observed values in the data set are above -1.00. Here the large probability values are quite large and decrease very slowly in the entire displayed range for K from

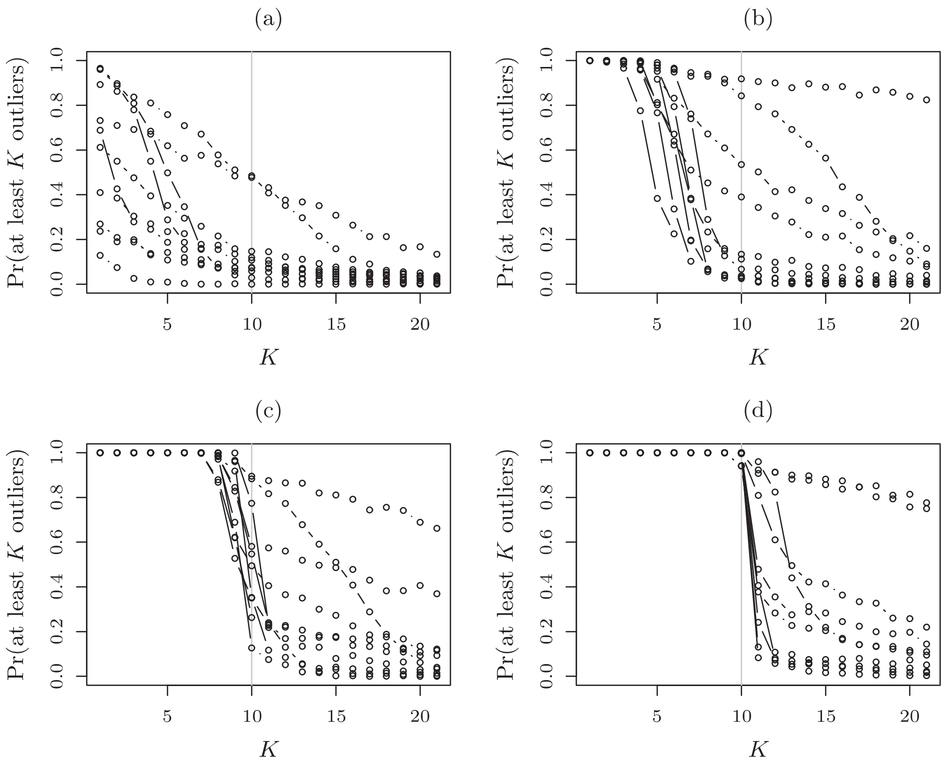

Figure 1. MB results for detecting outliers in 100 normal means. Each plot shows the posterior probability Pr (there are *at least* K outliers) given each of 10 replicates of simulated data based on generated normal means $\mu_1, \ldots, \mu_{100}$ having (a) no outliers; (b) 10 outliers generated from $2 + \text{Expo}(1)$; (c) 10 outliers generated from $4 + \text{Expo}(1)$; and (d) 10 outliers generated from $6 + \text{Expo}(1)$. The case in (b) with large probabilities in the displayed range corresponds to a simulated data having all the observed data values larger than -1.00.

1 to 21. This is not surprising because the MB analysis here tries to find a subset of data that consists of as many as possible data values under the condition that the subset *looks like* a typical sample from $N(0,1)$. This phenomenon can be seen for some cases in Figure 1(c) and (d), where MB would do a pretty good job for detecting outliers.

We note that finding the number of "outliers" is important in the context of multiple testing. The MB method provides a new approach to inference about the fraction of μ_i that are zero (see, e.g., Efron (2004)). We are currently investigating MB methods, including MB approaches to statistical deconvolution, for multiple testing.

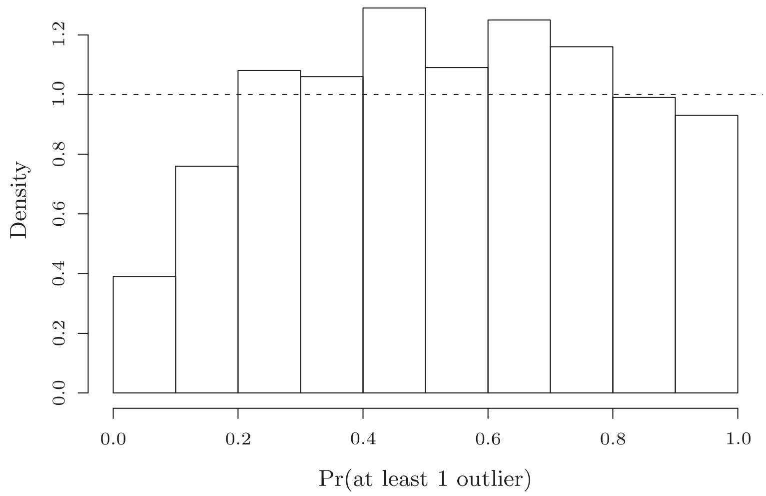

Figure 2. The histogram of the probability against the assertion that "there are no outliers ($\mu_i \neq 0$)" (or for the assertion that "there is at least one outlier") based on 1,000 simulated data sets from the model $X_i \overset{i.i.d.}{\sim} N(\mu_i, 1)$ with $\mu_i = 0$ for $i = 1, \ldots, n = 100$.

7. Discussion

For credible and efficient fiducial and DS parametric inference or building belief functions that have desired frequency properties, we have proposed WB and MB methods. Examples show that MB has the potential to resolve challenging statistical inference problems. The idea of WB can also be used to resolve non-uniqueness problems with DS (and fiducial) for a given sampling model. When a class of a-equations is under consideration, the fact that we "don't know" which a-equation is to be used would lead us to using WB models to capture the uncertainty about the choice of the a-equation.

We presented the work in the DS framework to build WB and MB models by modifying BDSMs. Nevertheless, WB and MB-DSMs are indeed pure DSMs, where the conditional DSMs for X given θ in the context of a-equation (1.1) should be interpreted as for situation-specific prediction rather than for data-generation. We plan to make more detailed argument for this view elsewhere. Also, more research is needed on defining efficient classes of weak beliefs from which MB at both belief level and assertion level can be sought.

Acknowledgement

The authors are grateful to Professor Arthur P. Dempster for sharing his deep understanding of scientific inference as well as for extensive comments and discussions, especially on BDSMs in Section 2 and WBs in Section 3. We thank

Diane Lambert, Earl Lawrence, David James, Ryan Martin, Jose Pinheiro, Donald B. Rubin, Glenn Shafer, and Scott Vander Wiel for helpful discussions, and the editorial board and two referees for constructive suggestions and critical comments on earlier versions of the paper. Jianchun Zhang's research is supported by the Purdue Research Foundation.

References

Almond, R. G. (1995). *Graphical Belief Modeling.* Chapman and Hall.

Brown, L. D., Cai, T. T. and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.* **16**, 101-117.

Bunke, H. (1975). Statistical inference: Fiducial and structural vs. likelihood. *Math. Operationsforsch. u. Statist.* **6**, 667-676.

Clopper, C. and Pearson, S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404-413.

Dawid, A. P. and Stone, M. (1982). The functional-model basis of fiducial inference. *Ann. Statist.* **10**, 1054-1067.

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

Dempster, A. P. (1967a). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.* **38**, 325-339.

Dempster, A. P. (1967b). Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika* **54**, 515-528.

Dempster, A. P. (1968a). Upper and lower probabilities generated by a random closed interval. *Ann. Math. Statist.* **39**, 957-966.

Dempster, A. P. (1968b). A generalization of Bayesian inference. *J. Roy. Statist. Soc. Ser. B* **30**, 205-247.

Dempster, A. P. (1969). Upper and lower probability inferences for families of hypothesis with monotone density ratios. *Ann. Math. Statist.* **40**, 953-969.

Dempster, A. P. (1990). Construction and local computation aspects of network belief functions. *Influence Diagrams, Belief Nets and Decision Analysis* (Edited by R. M. Oliver and J. Q. Smith), 121-141. John Wiley and Sons, Chichester.

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

Denoeux, T. (2006). Constructing belief functions from sample data using multinomial confidence regions. *Int. J. Approx. Reasoning* **42**, 228-252.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Amer. Statist. Assoc.* **99**, 96-104.

Fraser, D. A. S. (1966). Structural probability and a generalization. *Biometrika* **53**, 1-9.

Glover, F. (1967). Maximum matching in a convex bipartite graph. *Nav. Res. Logist. Q.* **14**, 313-316.

Hannig, J., (2006). On Fiducial Inference - the good, the bad and the ugly, Technical Report, Department of Statistics, Colorado State University.

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

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22**, 400-407.

Shafer, G. (1973). Allocations of probability: A theory of partial belief. Ph.D. thesis, Statistics, Princeton University. http://www.glennshafer.com/books/73dissertation.html

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

Shenoy, P. P. and Glenn Shafer, G. (1986). Propagating belief functions with local computations. *IEEE Expert* **1**, 43-52.

Soares, J. and Stefanes, M. A. (2007). Algorithms for maximum independent set in convex bipartite graphs. *Algorithmica* DOI 10.1007/s00453-007-9006-9.

Walley, P. (2002). Reconciling frequentist properties of with the likelihood principle. *J. Statist. Plann. Inference* **105**, 35-65.

Yager, R. R. and Liu, L. (2008). *Classic Works of the Dempster-Shafer Theory of Belief Functions.* Springer.

Department of Statistics, Purdue University, 250 N. University St., West Lafayette, IN, 47907-2067, USA.

E-mail: zhang10@stat.purdue.edu

Department of Statistics, Purdue University, 250 N. University St., West Lafayette, IN, 47907-2067, USA.

E-mail: chuanhai@stat.purdue.edu

(Received November 2007; accepted November 2009)