A note on Dempster-Shafer Recombination of Confidence Distributions

Jan Hannig*
Department of Statistics and Operations Research
The University of North Carolina at Chapel Hill

Min-ge Xie†
Department of Statistics and Biostatistics
Rutgers University

October 4, 2012

Abstract

It is often the case that there are several studies measuring the same parameter of interest. Naturally, it is of interest to provide a systematic way to combine the information from these studies. Examples of such situations include clinical trials, key comparison trials and other problems of practical importance. Singh *et al.* (2005) provide a compelling framework for combining information from multiple sources using the framework of confidence distributions. In this paper we investigate the feasibility of using the Dempster-Shafer recombination rule on this problem. We derive a practical combination rule and show that under assumption of asymptotic normality, the Dempster-Shafer combined confidence distribution is asymptotically equivalent to one of the method proposed in Singh *et al.* (2005). Numerical studies and comparisons for the common mean problem and the odds ratio in 2×2 tables are included.

*Jan Hannig's research was supported in part by the National Science Foundation under Grant No. 1007543 and 1016441.

†Minge Xie's research was supported in part by the National Science Foundation under Grant No. 0851521, 0915139 and 1107012.

1 Introduction

In many cases, there are multiple estimates of a particular quantity of interest arising from different experiments, or representing the particular quantity with respect to different populations, locations, years, etc. This scenario occurs often in, but not limited to, the context of meta-analysis, in which it is often desired to combine information from independent studies. There are numerous developments in this area. The range of meta-analysis approaches include the classical approaches of combining p-values, model-based (both fixed-effects and random effects) meta-analysis approaches, as well as specialized meta-analysis approaches targeting specific settings such as the Mantel-Haenszel method, Peto's method and a recently proposed exact meta-analysis approach by Tian *et al.* (2009) on combining confidence intervals, among others. Singh *et al.* (2005) proposed a simple but general recipe for combining *confidence distributions* from independent studies. Xie *et al.* (2011) and subsequent research showed that this general recipe and its extension can provide a unifying framework for almost all information combination methods used in current practice, including all aforementioned meta-analysis approaches. This unifying framework provides a compelling theoretical framework to understand and explore existing combining information approaches and also to develop new methodologies; cf., Xie *et al.* (2011).

Since confidence distributions are associated with Fisher's fiducial distributions and the Dempster-Shafer theory of belief functions, natural questions are whether it is possible to use the general Dempster-Shafer recombination rule (Dempster, 2008) to combine confidence distributions and how such a rule would relate to the general combination framework of Singh *et al.* (2005). This paper investigates these questions and in so doing links for the first time the seemingly unrelated research directions of Dempster-Shafer calculus and frequentist confidence distributions together through the combination.

The concept of confidence distributions has often been loosely referred to as a sample-dependent distribution function that can represent confidence intervals of all levels for a parameter of interest, see Cox (1958). It has a long history, especially when its interpretation is fused with fiducial inference (Fisher, 1973; Efron, 1993). Historically, it has been long considered as part of fiducial inference, although recent developments tend to define and interpret it within the frequentist framework without involving any fiducial reasoning, see Xie and Singh (2012) for a comprehensive review on the concept. In recent years, the notion of confidence distribution has attracted a surge of renewed attention. Together with the developments on generalized fiducial inferences and belief functions under Dempster-Shafer theory, it represents an emerging new research field to address problems where frequentist methods with good properties were previously unavailable.

In this paper we link together seemingly unrelated research directions of Dempster-Shafer calculus and confidence distributions. The paper is organized as follows. We first present in the remainder of this Introduction section the basic ideas of confidence distribution, Dempster-Shafer calculus, generalized fiducial inference and our extensions. Section 2 develops practical procedures for combining confidence distributions for either discrete or continuous data based on the Dempster-Shafer recombination rule. Section 3 shows that the combined confidence distributions are asymptotically equivalent to one of the methods proposed in Singh *et al.* (2005). Section 4 discusses results of two simulation studies, as well as provides real data examples and comparisons for the common mean problem and the odds ratio in 2×2 tables using two real data sets from the literature. Technical proofs are provided in Appendix.

In order to prevent confusion we use the following notation: Any object, e.g., density or centering, that is connected to the confidence distribution combined using the Dempster-Shafer recombination rule will have superscript $^{(DS)}$. Similarly any object connected to the confidence distribution combined using the rule of Singh *et al.* (2005) will have superscript $^{(c)}$.

1.1 Confidence Distributions

Let us assume that the observed data were generated from some distribution with parameters $(\theta_0, \xi_0) \subset \Theta \times \Xi$, where θ_0 is a one-dimensional parameter of interest and ξ_0 is a nuisance parameter. Denote by $\mathbf{X}$ the random sample, $\mathbf{x}$ its sample realization, and $\mathcal{X}$ the sample space. The following is a frequentist definition formulated by both Schweder and Hjort (2002) and Singh *et al.* (2001, 2005), where the parameters (θ_0, ξ_0) are treated as unknown fixed (not random) quantities.

Definition 1. A sample-dependent function $H(\cdot, \mathbf{x})$ on $\Theta \times \mathcal{X} \to [0, 1]$ is called a *confidence distribution (CD)* for the parameter of interest θ, if
(i) $H(\theta, \mathbf{x})$ is a cumulative distribution function in θ for a given sample $\mathbf{x}$ and (ii) $H(\theta_0, \mathbf{X})$ follows the standard uniform $U[0, 1]$ distribution under the

sample probability measure $P_{(\theta_0,\xi_0)}$.

The function $H(\theta, \mathbf{X})$ is an *asymptotic confidence distribution* if (ii) is true only asymptotically.

The density $h(\theta, \mathbf{x}) = (\partial/\partial\theta)H(\theta, \mathbf{x})$, if it exists, is called *a confidence density*, also known as *a CD density*; see, e.g., Efron (1993); Singh *et al.* (2007). For each fixed observed data $\mathbf{x}$, Singh *et al.* (2007) and Xie and Singh (2012) call a random variable $Q \in \Theta$ distributed according to the confidence distribution function $H(\theta, \mathbf{x})$ a *CD random variable*.

The concept of confidence distribution is quite broad, it encompasses and unifies a wide range of examples; from regular parametric cases to bootstrap distributions, p-value functions, normalized likelihood functions and, in some cases, Bayesian priors and Bayesian posteriors; see, e.g., Singh *et al.* (2005); Xie and Singh (2012). In particular, generalized fiducial distribution as described in Hannig *et al.* (2006); Hannig (2009, 2012); Wang *et al.* (2012) is often an asymptotic confidence distribution.

Singh *et al.* (2005) proposed a simple but general recipe for combining confidence distributions from, say k, independent studies, using a coordinate-wise monotonic function from a k-dimensional cube $[0,1]^k$ to the real line $\mathcal{R} = (-\infty, \infty)$. The recipe is an extension of the combining rules of the classical methods of combining p-values. Specifically, let $H_i(\cdot, \mathbf{x}_i)$ be a confidence distribution for θ from the sample $\mathbf{x}_i$ of the ith study and suppose $g_c(u_1, \ldots, u_k)$ is a given continuous function on $[0,1]^k \to \mathcal{R}^1$ which is nondecreasing in each coordinate. Singh *et al.* (2005) proposed a general framework for combining the k independent confidence distributions $H_i(\cdot, \mathbf{x}_i)$, $i = 1, \ldots, k$:

$$H^{(c)}(\theta, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) = G_c\{g_c(H_1(\theta, \mathbf{x}_1), \ldots, H_k(\theta, \mathbf{x}_k))\}. \qquad (1)$$

where the function $G_c(\cdot)$ is completely determined by the monotonic g_c function with $G_c(t) = P(g_c(U_1, \ldots, U_k) \le t)$. Here, $U_1, \ldots, U_k$ being independent $U[0,1]$ random variables. The combined function $H^{(c)}(\cdot)$ contains information from all k studies and Singh *et al.* (2005) showed that the combined function $H^{(c)}(\cdot)$ is a confidence distribution for the parameter θ.

A nice feature of the combining method (1) is that it does not require any information regarding how the input confidence distributions $H_i(\cdot)$ are obtained, aside from the assumed independence. Xie *et al.* (2011) and subsequent research showed that this general recipe and its extension can provide a unifying framework for almost all information combination methods used in current practice. This includes the classical approaches of combining p-values, e.g., Fisher, Stouffer, Tippett, Max and Sum methods, and the modern model-based meta-analysis approach, e.g., fixed and random effects models.

A special class of choices for g_c illustrated by Singh *et al.* (2005) is:

$$g_c(u_1, \ldots, u_k) = F^{-1}(u_1) + \ldots + F^{-1}(u_k),$$

where $F(\cdot)$ is a given cumulative distribution function. In this case, $G_c(\cdot)$ is the convolution of the k copies of $F(\cdot)$. When $F(t) = \exp(t)$, for $t < 0$, is the cumulative distribution function of the negative exponential distribution, the recipe (1) is an extension of Fisher's way of combining p-values. When $F(t) = \Phi(t)$, the cumulative distribution function of the standard normal, we have

$$H^{(c)}(\theta, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) = \Phi\left(\frac{1}{\sqrt{k}}\sum_{i=1}^{k} \Phi^{-1}\{H_i(\theta, \mathbf{x}_i)\}\right), \tag{2}$$

which is an extension of Stouffer's way of combining p-values. Xie *et al.* (2011) suggested to include weights in the combination to improve efficiency. In particular, (2.2) of Xie *et al.* (2011) suggested using

$$H^{(c)}(\theta, \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) = \Phi\left(\frac{\sum_{i=1}^{k} w_i^{-1}\Phi^{-1}\{H_i(\theta, \mathbf{x}_i)\}}{(\sum_{i=1}^{k} w_i^{-2})^{1/2}}\right), \tag{3}$$

where the weights w_i could be sample dependent. In Section 4 of this article, we focus on the special combination rule (3) with $w_i = \tau_i \stackrel{\text{def}}{=} \{H_i^{-1}(.75) - H_i^{-1}(.25)\}/\{2\Phi^{-1}(0.75)\}$, where $H_i^{-1}(\beta)$ is the β quantile of the confidence distribution $H_i(\theta, \mathbf{x}_i)$ (i.e., it solves the θ equation $H_i(\theta, \mathbf{X}_i) = \beta$ for a given $0 \leq \beta \leq 1$). In this case (with sample-dependent weights and under mild conditions), inference based on (3) is only asymptotically valid.

In Section 1.2 next, we provide an introduction of the Dempster-Shafer theory of inference and demonstrate that confidence distributions as defined in Definition 1 also fit into this framework. This relationship allows us to derive an alternative approach for combining confidence distributions.

1.2 Dempster-Shafer theory

In this section we provide a brief introduction of the Dempster-Shafer theory. A more thorough introduction can be found in Dempster (2008) and Shafer (1976). Some comments can be also found in Hannig (2009) and Zhang and Liu (2011). The main purpose of the Dempster-Shafer (DS) theory is to convert observed data and pivotal relationships to "upper" and "lower" probability statements. Mathematically, these statement are derived with the help of random subsets of the parameter space.

In particular, Dempster starts with an auxiliary equation

$$0 = a(\mathbf{X}, \theta, \mathbf{U})$$

relating the observable data vector $\mathbf{X} \in \mathcal{X}$, the parameters $\theta \in \Theta$ and an auxiliary random vector $\mathbf{U}$ with a fully known distribution independent of any parameters, e.g., vector of independent standard uniforms $U(0,1)$. Traditionally, the auxiliary equation is either in the form of a data generating equation

$$\mathbf{X} = G(\theta, \mathbf{U}) \tag{4}$$

or a pivotal equation $\mathbf{U} = H(\mathbf{X}, \theta)$ and is chosen based on the distribution of the observable data.

DS theory then inverts the auxiliary equation into the multivalued mapping

$$\mathcal{M}(\mathbf{U}) = \{(\mathbf{X}, \theta),\ a(\mathbf{X}, \theta, \mathbf{U}) = 0\}$$

that is called the DS model. After observing a particular data vector $\mathbf{x}$ we constrain the multivalued mapping to a random set of parameters

$$\mathcal{M}_{\mathbf{x}}(\mathbf{U}) = \{\theta,\ a(\mathbf{x}, \theta, \mathbf{U}) = 0\}.$$

For any assertion $A \subset \Theta$ about the parameters the DS theory then gives three probabilities

$$p = \frac{P(A \subset M_{\mathbf{x}}(\mathbf{U}))}{P(M_{\mathbf{x}}(\mathbf{U}) \neq \emptyset)}, \quad q = \frac{P(A^{\complement} \subset M_{\mathbf{x}}(\mathbf{U}))}{P(M_{\mathbf{x}}(\mathbf{U}) \neq \emptyset)}, \quad r = 1 - p - q. \tag{5}$$

Here p is interpreted as the probability in support of A, q the probability in contradiction to A and r the probability "do not know", supporting neither A nor $A^{\complement}$.

By inspecting (5) we see that DS inference is based on random set $\tilde{Q}$ with a distribution given by the conditional distribution of

$$\tilde{Q} \sim \left[\mathcal{M}_{\mathbf{x}}(\mathbf{U}) \mid \{\mathcal{M}_{\mathbf{x}}(\mathbf{U}) \neq \emptyset\}\right], \tag{6}$$

so that $p = P(\tilde{Q} \subset A)$ and $q = P(\tilde{Q} \subset A^{\complement})$. In this article we will call $\tilde{Q}$ *belief random set*.

Let us demonstrate this on two simple examples. First, consider the simple example of a single observation x from the $N(\theta, 1)$. The appropriate data generating equation is $X = \theta + U$, where $U \sim N(0, 1)$. After observing x the constraint multivalued mapping is the singleton $\mathcal{M}_x(U) = \{x - U\}$ and $\tilde{Q} = \{Q\}$ where Q follows $N(x, 1)$.

Second, let $X_1, \ldots, X_n$ be a sample from Bernoulli(p). A possible data generating equation is $X_i = I_{(0,p)}(U_i), i = 1 \ldots, n$, with U_i i.i.d. $U(0, 1)$. After observing the vector $\mathbf{x}$ the constraint multivalued mapping

$$\mathcal{M}_{\mathbf{x}}(\mathbf{U}) = \left\{ p \in [0, 1] : \quad \begin{matrix} p \leq U_i \text{ if } x_i = 0 \\ p \geq U_i \text{ if } x_i = 1 \end{matrix} \right\}.$$

An exchangeability argument shows that the belief random set has the same distribution as the random interval interval $\tilde{Q} = [U_{(x)}, U_{(x+1)}]$ where $U_{(x)}$ is the xth order statistics of $U_1, \ldots, U_n$.

Dempster-Shafer theory provides a recombination rule to combine information from several sources into a single object. We will state this recombination rule in the language of belief random sets. Let $\tilde{Q}_1, \ldots, \tilde{Q}_k$ be belief random sets to be combined. The combined belief random set $\tilde{Q}$ will have as its distribution the following conditional distribution

$$\bigcap_{i=1}^{k} \tilde{Q}_i \mid \{\bigcap_{i=1}^{k} \tilde{Q}_i \neq \emptyset\}. \tag{7}$$

Both (6) and (7) provide a general recipe. However, it is not always clear how to implement them in any given particular situation. In particular the conditional distribution in (6) and (7) is not uniquely defined if the condition has probability zero due to the Borel paradox (Casella and Berger, 2002). For this reason, Dempster-Shafer theory is predominantly applied to discrete distributions where the problem of the Borel paradox does not arise. In the next section we comment on how this limitation can be overcome using ideas of generalized fiducial inference.

Confidence distributions can be formally put into the DS framework as follows. Let $H(\theta, \mathbf{X})$ be a confidence distribution function. The equation $H(\theta, \mathbf{X}) = U$ is a pivotal equation, since by Definition 1 the random variable U has the standard uniform distribution $U(0, 1)$. Assuming that the solution to $H(\cdot, \mathbf{x}) = u$ exists for the observed $\mathbf{x}$ and almost all $u \in (0, 1)$, the belief random set defined in (6) is a singleton $\tilde{Q} = \{Q\}$ where Q is a CD random variable distributed according to $H(\cdot, \mathbf{x})$.

1.3 Generalized Fiducial Inference

The aim of generalized fiducial inference is to define a measure on the probability space by inversion from the structural generating equation (4). In this section we explain two ideas of generalized fiducial inference pertinent to this paper.

In the case when the belief random set is a singleton $\tilde{Q} = \{Q\}$ with probability 1, the generalized fiducial distribution is the same as the distribution of Q. When the belief random set is not a singleton Hannig (2009) suggest selecting one of the elements of $\tilde{Q}$ based on a predetermined, possibly random rule. Hannig (2012) shows that for many popular models the effects of this choice disappear asymptotically.

If the belief random set is an interval $\tilde{Q} = [Q^-, Q^+]$ with probability 1 then it is often recommended to maximize the variance of the fiducial distribution by selecting either end of the interval with probability 0.5. This selection is called "half correction" (Efron, 1998; Schweder and Hjort, 2002; Hannig, 2009) and we will use it in Section 2.2.

A more serious issue arises when $P(\mathcal{M}_x(U) \neq \emptyset) = 0$. In this case the quantities in (5) and (6) are not well defined due to Borel paradox. Hannig (2012) recommends a plausible resolution of this non-uniqueness by discretizing the data.

In particular, define

$$\mathcal{M}_{x,\epsilon}(U) = \{\theta,\ \|x - G(\theta, U)\|_\infty < \epsilon\}.$$

Here $\|v\|_\infty$ is the l^∞ norm of the vector v. Notice that for any observable x and $\epsilon > 0$ the $P(\mathcal{M}_{x,\epsilon}(U) \neq \emptyset) > 0$. The generalized fiducial distribution of Hannig (2012) is the weak limit of the conditional distributions

$$\tilde{Q} \sim \lim_{\varepsilon \to 0} \left[\mathcal{M}_{x,\epsilon}(U) \mid \{\mathcal{M}_{x,\epsilon}(U) \neq \emptyset\}\right].$$

Under weak assumptions Hannig (2012) shows the limit exists and provides a formula for the density of the generalized fiducial distribution. For example, if $\theta \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$ and the inverse to (4) denoted by $\mathbf{u} = H(\mathbf{x}, \theta)$

exists, the density of generalized fiducial distribution simplifies to

$$r(\theta) \propto J(\mathbf{x}, \theta) f(\mathbf{x}, \theta),$$

where $f(\mathbf{x}, \theta)$ is the likelihood and

$$J(\mathbf{x}, \theta) = \left\| \frac{\partial}{\partial \theta} G(\theta, \mathbf{u}) \Big|_{\mathbf{u}=H(\mathbf{x},\theta)} \right\|_1$$

where $\|v\|_1$ is the l^1 norm of the vector v.

In Section 2.1 we will use the same discretization idea to resolve non-uniqueness due to the Borel Paradox in the DS recombination rule (7).

2 Dempster-Shafer recombination rules

In this section we derive Dempster-Shafer based formulas for combining confidence distributions derived from either discrete data or continuous data. Continuous data are dealt with in Section 2.1 and discrete data are addressed in Section 2.2.

2.1 Dempster-Shafer recombination rule for continuous data

Let us assume that we have k confidence distributions for a single parameter θ based on independent data sets. In particular we assume that we have $H_i(\theta, \mathbf{x}_i)$, $i = 1, \ldots, k$, each satisfying conditions (i) and (ii) of Definition 1.

In this section we will make the following additional assumptions for all $i = 1, \ldots, k$ and an open $A \in \theta$

(A1) For all u_i and all $\mathbf{x}$ in a neighborhood of $\mathbf{x}_i$ the equation $H_i(\cdot, \mathbf{x}) = u_i$ has a unique solution $Q_i(u_i, \mathbf{x})$.

(A2) The partial derivatives $(\partial/\partial\theta)H_i(\theta, \mathbf{x})$ and the gradients $\nabla_{\mathbf{x}} H_i(\theta, \mathbf{x})$ are continuous for all θ and all $\mathbf{x}$ in a neighborhood of $\mathbf{x}_i$.

(A3) For all $\theta \in A$, the Euclidean norm (l^2) of the gradient

$$D_{\mathbf{x}_i} H_i(\theta, \mathbf{x}_i) = \|\nabla_{\mathbf{x}} H_i(\theta, \mathbf{x}_i)\|_2 > 0. \tag{8}$$

We remark that the conditions (A1) – (A3) are well suited for confidence distributions derived from continuously distributed data sets. They are satisfied for most usual continuous distributions such as exponential family. Discretely distributed data will be dealt with in the next section.

The belief random sets are $\tilde{Q}_i = \{Q_i(U_i, \mathbf{x}_i)\}$. Consequently, the Dempster-Shafer recombination rule (7) gives the distribution of the combined belief random set as

$$\{Q_i(U_i, \mathbf{x}_i)\} \mid Q_1(U_1, \mathbf{x}_1) = \cdots = Q_k(U_k, \mathbf{x}_k), \tag{9}$$

where U_i are i.i.d. U(0,1). Unfortunately the condition in (9) has probability 0 and therefore the conditional distribution in (9) is not unique due to the Borel paradox.

We follow the spirit of the generalized fiducial inference in interpreting (9). Using the Euclidean neighborhoods of $\mathbf{x}_i$ denote

$$\tilde{Q}_{i,\epsilon}(u_i, \mathbf{x}_i) = \{\theta : H_i(\theta, \mathbf{x}) = u_i \text{ for some } \|\mathbf{x} - \mathbf{x}_i\|_2 < \epsilon\}.$$

Then define the distribution of the DS recombined belief random set as the weak limit as $\epsilon \to 0$ of conditional distributions

$$\bigcap_{i=1}^{k} \tilde{Q}_{i,\epsilon}(U_i, \mathbf{x}_i) \mid \{\bigcap_{i=1}^{k} \tilde{Q}_{i,\epsilon}(U_i, \mathbf{x}_i) \neq \emptyset\}.$$

Continuity implies that $Q_{i,\epsilon}(u_i, \mathbf{x}_i) = [Q^-_{i,\epsilon}(u_i, \mathbf{x}_i), Q^+_{i,\epsilon}(u_i, \mathbf{x}_i)]$. The existence of the total derivative, Cauchy-Schwartz inequality and some calculus shows that the limiting belief random set $\tilde{Q}^{(DS)} = \{Q^{(DS)}\}$, where $Q^{(DS)}$ is a random variable with density

$$h^{(DS)}(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k) \propto \mathcal{L}(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k) J(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k). \tag{10}$$

Here

$$\mathcal{L}(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k) = \prod_{i=1}^{k} D_{\mathbf{x}_i} H_i(\theta, \mathbf{x}_i),$$

with $D_{\mathbf{x}_i} H_i(\theta, \mathbf{x}_i)$ defined in (8), could be viewed as a profile likelihood induced by the confidence distribution and

$$J(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k) = \sum_{i=1}^{k} \frac{\left|\frac{\partial}{\partial\theta} H_i(\theta, \mathbf{x}_i)\right|}{D_{\mathbf{x}_i} H_i(\theta, \mathbf{x}_i)}$$

is similar to the fraction of Jacobians seen in Hannig (2012).

To demonstrate this rule we will consider two examples.

Example 1. Let us consider k independent normal samples with common unknown mean and known variances. Denote the standard normal density and distribution function by $\varphi(z)$ and $\Phi(z)$ respectively. The individual confidence distributions based on each of the k samples are

$$H_i(\mu, \mathbf{x}_i) = \Phi\left(\frac{\mu - \bar{x}_i}{\sigma_i/\sqrt{n_i}}\right), \quad i = 1, \dots k,$$

respectively. The Dempster-Shafer recombined density (10) is proportional to

$$h^{(DS)}(\mu|\bar{x}_1, \dots, \bar{x}_k) \propto \prod_{i=1}^{k} \varphi\left(\frac{\mu - \bar{x}_i}{\sigma_i/\sqrt{n_i}}\right).$$

A simple calculation shows that the recombined confidence distribution is the normal distribution with mean $(\sum_{i=1}^{k} \sigma_i^{-2} n_i \bar{x}_i)/(\sum_{i=1}^{k} \sigma_i^{-2} n_i)$ and variance $(\sum_{i=1}^{k} \sigma_i^{-2} n_i)^{-1}$. This is the same as the combined confidence distribution using (3).

Example 2. Let us consider k independent normal samples with common unknown mean and unknown unequal variances. Denote the density and distribution function of the t distribution with m degrees of freedom by $f_m(z)$ and $F_m(z)$ respectively. The individual confidence distributions are

$$H_i(\mu, \mathbf{x}_i) = F_{n_i - 1}\left(\frac{\mu - \bar{x}_i}{s_i/\sqrt{n_i}}\right), \quad i = 1, \dots k,$$

respectively. Here $\bar{x}_i$ and s_i are the sample mean and standard deviation of the ith sample.

To compute the Dempster-Shafer recombined density (10) we evaluate

$$\frac{\partial}{\partial \mu} H_i(\mu, \mathbf{x}_i) = f_{n_i - 1}\left(\frac{\mu - \bar{x}_i}{s_i/\sqrt{n_i}}\right) \frac{\sqrt{n_i}}{s_i} = \frac{\sqrt{n_i} C_{n_i}}{s_i} \left\{1 + \frac{n_i(\mu - \bar{x}_i)^2}{(n_i - 1)s_i^2}\right\}^{-n_i/2}$$

and

$$\|\nabla_{\mathbf{x}_i} H_i(\mu, \mathbf{x}_i)\|_2 = f_{n_i - 1}\left(\frac{\mu - \bar{x}_i}{s_i/\sqrt{n_i}}\right) \left\|\nabla_{\mathbf{x}_i} \frac{\mu - \bar{x}_i}{s_i/\sqrt{n_i}}\right\|_2 = \frac{C_{n_i}}{s_i} \left\{1 + \frac{n_i(\mu - \bar{x}_i)^2}{(n_i - 1)s_i^2}\right\}^{-(n_i - 1)/2}.$$

Thus the DS combined density is

$$h^{(DS)}(\mu|\mathbf{x}_1,\ldots,\mathbf{x}_k) \propto \left[\sum_{i=1}^{n}\left\{\frac{1}{n_i}+\frac{(\mu-\bar{x}_i)^2}{(n_i-1)s_i^2}\right\}^{-1/2}\right]\prod_{i=1}^{k}\left\{1+\frac{n_i(\mu-\bar{x}_i)^2}{(n_i-1)s_i^2}\right\}^{-(n_i-1)/2}. \tag{11}$$

Surprisingly, this is not the same as the generalized fiducial distribution for the common mean problem using the pooled data (Hannig *et al.*, 2006). Regardless, arguments similar to the arguments in Hannig *et al.* (2006) prove (11) is an asymptotic confidence distribution.

We study small sample performance of (11) in Section 4.1. In particular we compare (11) with the confidence distributions combined using methods (2) and (3) in terms of coverage and median length of 95% confidence intervals. A numerical example in Section 4.2 illustrates the use of this recombined distribution on a real data example.

2.2 Dempster-Shafer recombination rule for discrete data

When dealing with confidence distributions for discrete data, it is often the case that there is a range of acceptable (approximate) confidence distributions; this is due to the discrete nature of the data. This uncertainty due to discretization is often dealt with by splitting the difference and applying the "half correction" (Efron, 1998; Schweder and Hjort, 2002; Hannig, 2009). Here, "acceptable" means that these distributions can be utilized to make valid inference, e.g., they are asymptotic confidence distributions.

Let us assume that we have $H_i(\theta,\mathbf{x}_i)$, $i=1,\ldots,k$, where the "half corrected" CD has been obtained by averaging the right and left limit $\{H_i^+(\theta,\mathbf{x}_i)+H_i^-(\theta,\mathbf{x}_i)\}/2$, where both $H_i^{\pm}(\theta,\mathbf{x}_i)$ are approximate CDs, see Definition 1. By the nature of "half correction", any distribution function between $H_i^+(\theta,\mathbf{x}_i)$ and $H_i^-(\theta,\mathbf{x}_i)$ is an approximate confidence distribution and so we define the belief random sets as

$$\tilde{Q}_i(u,\mathbf{x}_i)=\{\theta : u\in[H_i^+(\theta,\mathbf{x}_i),H_i^-(\theta,\mathbf{x}_i)]\}$$

with the understanding that if $a>b$ then we reverse the interval $[a,b]$ to $[b,a]$.

The Dempster-Shafer recombination rule (7) simplifies in this setting to the conditional distribution of

$$\bigcap_{i=1}^{k} \tilde{Q}_i(U_i, \mathbf{x}_i) \mid \{\bigcap_{i=1}^{k} \tilde{Q}_i(U_i, \mathbf{x}_i) \neq \emptyset\}, \tag{12}$$

where U_i are i.i.d. U(0,1).

To simplify (12) into a workable formula we make the following assumptions for all $i = 1, \ldots, k$ and an open $A \subset \Theta$.

(A1') For all u_i and the observed $\mathbf{x}_i$ the equations $H_i^+(\cdot, \mathbf{x}_i) = u_i$ and $H_i^-(\cdot, \mathbf{x}_i) = u_i$ have a unique solution $Q_i^+(u_i, \mathbf{x}_i)$ and $Q_i^-(u_i, \mathbf{x}_i)$ respectively.

(A2') The partial derivatives $(\partial/\partial\theta) H_i^+(\theta, \mathbf{x}_i)$ and $(\partial/\partial\theta) H_i^-(\theta, \mathbf{x}_i)$ are continuous for all θ.

(A3') For all $\theta \in A$, the absolute value of the difference

$$D_{\mathbf{x}_i} H_i(\theta, \mathbf{x}_i) = \left| H_i^-(\theta, \mathbf{x}_i) - H_i^+(\theta, \mathbf{x}_i) \right| > 0. \tag{13}$$

These assumptions are satisfied for the usual discrete families of distributions such as those based on exponential families.

Notice that the assumptions imply $\tilde{Q}_i(U_i, \mathbf{x}_i) = [Q_i^-(U_i, \mathbf{x}_i), Q_i^+(U_i, \mathbf{x}_i)]$ modulo a possible reversal of the interval. Consequently the intersection $\bigcap_{i=1}^{k} \tilde{Q}_i(U_i, \mathbf{x}_i)$ is an interval or an empty set. A simple calculation shows that the conditional distribution (12) is a random interval $[Q^{(DS)-}, Q^{(DS)+}]$, where $Q^{(DS)\pm}$ have a marginal density proportional to

$$h^{(DS)\pm}(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k) \propto \mathcal{L}(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k) J^{\pm}(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k)$$

respectively, with

$$\mathcal{L}(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k) = \prod_{i=1}^{k} D_{\mathbf{x}_i} H_i(\theta, \mathbf{x}_i),$$

$D_{\mathbf{x}_i} H_i(\theta, \mathbf{x}_i)$ defined in (13),

$$J^+(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k) = \sum_{i=1}^{k} \left| \frac{\frac{\partial}{\partial\theta} H_i^+(\theta, \mathbf{x}_i)}{D_{\mathbf{x}_i} H_i(\theta, \mathbf{x}_i)} \right| \quad \text{and} \quad J^-(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k) = \sum_{i=1}^{k} \left| \frac{\frac{\partial}{\partial\theta} H_i^-(\theta, \mathbf{x}_i)}{D_{\mathbf{x}_i} H_i(\theta, \mathbf{x}_i)} \right|$$

respectively. The half corrected Dempster-Shafer recombined density is then

$$h^{(DS)}(\theta|\mathbf{x}_1, \ldots, \mathbf{x}_k) = \frac{1}{2} \left\{ h^{(DS)-}(\theta) + h^{(DS)+}(\theta) \right\}. \tag{14}$$

We will demonstrate (14) on the following examples.

Example 3. Let us assume that we have k independent binomial samples with common probability of success p and number of trials $n_1, \ldots, n_k$. Denote the observed values by $x_1, \ldots, x_k$. The half corrected confidence distribution (Efron, 1998; Schweder and Hjort, 2002; Hannig, 2009) is obtained from $H_i(p, x_i) = P_p(X_i > x_i)$ as

$$\frac{H_i(p,x_i) + H_i(p,x_i-1)}{2} = \sum_{x_i<k\leq n_i} \binom{n_i}{k} p^k(1-p)^{n_i-k} + \frac{1}{2}\binom{n_i}{x_i} p^{x_i}(1-p)^{n_i-x_i}.$$

Thus $H_i^+(p,x_i) = H_i(p,x_i)$ and $H_i^-(p,x_i) = H_i(p,x_i-1)$. Notice that the half corrected confidence distribution function is the 50-50 mixture of the $\text{Beta}(x_i, n_i - x_i + 1)$ and $\text{Beta}(x_i + 1, n_i - x_i)$ distributions.

In order to evaluate (14) set $x = \sum_{i=1}^k x_i$ and $n = \sum_{i=1}^k n_i$ and compute

$$\begin{aligned}
\frac{\partial}{\partial p} H_i^+(p,x_i) &= \frac{p^{x_i}(1-p)^{n_i-x_i-1}}{B(x_i+1, n_i-x_i)}, \\
\frac{\partial}{\partial p} H_i^-(p,x_i) &= \frac{p^{x_i-1}(1-p)^{n_i-x_i}}{B(x_i, n_i-x_i+1)}, \\
D_{x_i} H_i(p,x_i) &= \binom{n_i}{x_i} p^{x_i}(1-p)^{n_i-x_i}.
\end{aligned}$$

From here we see immediately that $h^{(DS)+}(\theta|x_1,\ldots,x_k)$ and $h^{(DS)-}(\theta|x_1,\ldots,x_k)$ are the density of $\text{Beta}(x, n-x-1)$ and $\text{Beta}(x-1, n-x)$ respectively. Therefore the half corrected Dempster-Shafer recombined confidence distribution function is given by

$$\sum_{x<k\leq n} \binom{n}{k} p^k(1-p)^{n-k} + \frac{1}{2}\binom{n}{x} p^x(1-p)^{n-x}$$

which is the same as the half recombined confidence distribution computed from the pooled sample which is known to have good small and large sample properties (Hannig, 2009).

Example 4. In the meta-analysis literature, many articles have considered the setting of performing a combined inference for the common odds ratio across a series of 2×2 tables from binomial clinical trials; see Liu *et al.* (2011), and references therein. Consider a series of k independent 2×2 tables formed by binomial random variables (X_{1i}, X_{0i}) with probability of a success $(p_{1i}, p_{0i}$ and number of trials (n_{1i}, n_{0i}). Assume that the odds ratio

$$\psi = \frac{p_{1i}/(1-p_{1i})}{p_{0i}/(1-p_{0i})}$$

remains constant across the tables. Set $T_i = X_{1i} + X_{0i}$ and define

$$H_i(\psi, x_i, t_i) = P_\psi(X_i > x_i | T_i = t_i).$$

We have $H_i^+(\psi, x_i, t_i) = H_i(\psi, x_i, t_i)$, $H_i^-(\psi, x_i, t_i) = H_i(\psi, x_i - 1, t_i)$ and

$$D_x H_i(\psi, x, t_i) = P_\psi(X_i = x | T_i = t_i) = \frac{\binom{n_{1i}}{x}\binom{n_{0i}}{t_i - x}\psi^x}{\sum_{k=L_i}^{U_i}\binom{n_{1i}}{k}\binom{n_{0i}}{t_i-k}\psi^k}, \quad (x = L_i, \ldots, U_i).$$

with $L_i = \max(0, t_i - n_{0i})$ and $U_i = \min(n_{1i}, t_i)$.

To simplify the formulas set $F(\psi, n, m, t) = \sum_{k=L}^{U}\binom{n}{k}\binom{m}{t-k}\psi^k$. This is a constant multiple of the hypergeometric $2F1$ function. Notice that the derivative $F'(\psi, n, m, t) = nF(\psi, n-1, m, t-1)$. By collecting the terms together we obtain

$$h^{(DS)\pm}(\psi) \propto \frac{\psi^{\sum_{i=1}^k x_i - 1} J^{\pm}(\psi)}{\prod_{i=1}^k F(\psi, n_{1i}, n_{0i}, t_i)} \tag{15}$$

where

$$J^+(\psi) = \sum_{i=1}^k \left| \sum_{l=1}^{U_i - x_i} \frac{\psi^l \binom{n_{1i}}{x_i + l}\binom{n_{0i}}{t_i - x_i - l}}{\binom{n_{1i}}{x_i}\binom{n_{0i}}{t_i - x_i}} \left((x_i + l) - \frac{n_{1i}\psi F(\psi, n_{1i} - 1, n_{0i}, t_i - 1)}{F(\psi, n_{1i}, n_{0i}, t_i)} \right) \right|$$

and

$$J^-(\psi) = \sum_{i=1}^k \left| \sum_{l=0}^{U_i - x_i} \frac{\psi^l \binom{n_{1i}}{x_i + l}\binom{n_{0i}}{t_i - x_i - l}}{\binom{n_{1i}}{x_i}\binom{n_{0i}}{t_i - x_i}} \left((x_i + l) - \frac{n_{1i}\psi F(\psi, n_{1i} - 1, n_{0i}, t_i - 1)}{F(\psi, n_{1i}, n_{0i}, t_i)} \right) \right|.$$

If desired, the density of the confidence distribution for the log odds ratio $\theta = \log\psi$ can be obtained by a simple change of variable.

We report results of a simulation study in Section 4.1 comparing confidence distribution combined using (15) with the confidence distributions combined using (3) in terms of coverage and median length of 95% confidence intervals. A numerical example in Section 4.2 illustrates the use of this recombined distribution on a real data example.

3 Asymptotic results

It is often the case that each of the confidence distributions we are recombining is asymptotically normal; i.e. $H_i(\theta, \mathbf{x}_i) \approx \Phi\left(\frac{\theta - T_i(\mathbf{x}_i)}{c_i}\right)$ for some statistic $T_i(\mathbf{x}_i)$ and scaling c_i. We explain in what sense the confidence distribution should be asymptotically close to normal in the assumptions below. We show that under the assumptions the confidence distribution combined using the Dempster-Shafer rule (10) is asymptotically equivalent to the combination rule (3). Additionally, we show that under the assumptions the Dempster-Shafer recombined distribution is an asymptotic confidence distribution.

In the assumptions below we assume that we have a sequence of asymptotic confidence distributions $H_{i,n}(\theta, \mathbf{X}_{i,n})$ each based on a sample $\mathbf{X}_{i,n}$ generated from a distributions with a common parameter of interest θ_0 and increasing sample size (see Assumption 1a). We also assume that these sample sizes grow to infinity at the same rate (see Assumption 3). To highlight the dependence of certain terms on sample sizes, we add subscript $_n$' to these terms whenever it applies in this section. For instance, in this section and in the Appendix we write $\mathbf{X}_{i,n}$ instead of $\mathbf{X}_i$, $H_{i,n}(\theta, \mathbf{X}_{i,n})$ instead of $H_i(\theta, \mathbf{X}_i)$, etc.

The assumptions and the theorem are formulated for the continuous case. In the discrete case we can modify the the Assumption 2a by slightly modifying Assumption 2b so that it holds for both $(\partial/\partial\theta)H^+_{i,n}(\theta, \mathbf{x}_{i,n})$ and $(\partial/\partial\theta)H^-_{i,n}(\theta, \mathbf{x}_{i,n})$. Then (16) in the Theorem 1 will hold for both $h_n^{(DS)\pm}$.

Assumption 1. For all $i = 1 \ldots k$ as $n \to \infty$:

(a) $H_{i,n}(\theta_0, \mathbf{X}_{i,n}) \xrightarrow{\mathcal{D}} U(0,1)$ and $\mathbf{X}_{1,n}, \ldots, \mathbf{X}_{k,n}$ are independent for each n.

(b) $H_{i,n}(\theta_0, \mathbf{X}_{i,n}) - \Phi\left(\frac{\theta_0 - t_{i,n}(\mathbf{X}_{i,n})}{c_{i,n}}\right) \xrightarrow{P} 0.$

Assumption 2. For all $i = 1, \ldots, k$ as $n \to \infty$

(a) $\int_\Theta \left| D_{\mathbf{x}_{i,n}} H_{i,n}(\theta, \mathbf{X}_{i,n}) - \frac{1}{c_{i,n}} \varphi\left(\frac{\theta - t_{i,n}(\mathbf{X}_{i,n})}{c_{i,n}}\right) \right| d\theta \xrightarrow{P} 0.$

(b) $\int_\Theta \left| \frac{\partial}{\partial\theta} H_{i,n}(\theta, \mathbf{X}_{i,n}) - \frac{1}{c_{i,n}} \varphi\left(\frac{\theta - t_{i,n}(\mathbf{X}_{i,n})}{c_{i,n}}\right) \right| d\theta \xrightarrow{P} 0.$

(c) $c_{i,n} D_{\mathbf{x}_{i,n}} H_{i,n}(\theta, \mathbf{X}_{i,n})$ is bounded in probability.

Assumption 3. For all $i = 1, \ldots, k$, $c_{i,n} \left(\sum_{j=1}^{k} c_{j,n}^{-2} \right)^{1/2} \to r_i \in (0, \infty)$ as $n \to \infty$.

We will first state a theorem showing asymptotic normality of the DS combined confidence distribution.

Theorem 1. *Suppose Assumptions 1, 2, 3. Using $t_{j,n}(\mathbf{x})$ and c_j defined in Assumption 2 define the centering*

$$T_n^{(DS)} = \frac{\sum_{j=1}^{k} t_{j,n}(X_{j,n}) c_{j,n}^{-2}}{\sum_{j=1}^{k} c_{j,n}^{-2}}.$$

Also denote by $\tilde{h}_n(\theta|t)$ the density of $N(t, \sum_{j=1}^{k} c_{j,n}^{-2})$. Then

$$\int_{\Theta} \left| h_n^{(DS)}(\theta|\mathbf{X}_{1,n}, \cdots, \mathbf{X}_{k,n}) - \tilde{h}_n(\theta|T_n^{(DS)}) \right| d\theta \xrightarrow{P} 0 \tag{16}$$

and the Dempster-Shafer recombined confidence distribution is an asymptotic confidence distribution.

The combined confidence distribution using the Dempster-Shafer recombined rule is

$$H_n^{(DS)}(\theta, \mathbf{X}_{1,n}, \cdots, \mathbf{X}_{k,n}) = \int_{-\infty}^{\theta} h_n^{(DS)}(\eta|\mathbf{X}_{1,n}, \cdots, \mathbf{X}_{k,n}) d\eta.$$

The following theorem states that, under certain conditions, the confidence distribution $H_n^{(c)}(\theta, \mathbf{X}_{1,n}, \mathbf{X}_{2,n}, \ldots, \mathbf{X}_{k,n})$ obtained by the CD combination recipe (3) is asymptotically equivalent to the recombination confidence distribution $H_n^{(DS)}(\theta, \mathbf{X}_{1,n}, \cdots, \mathbf{X}_{k,n})$ obtained by the Dempster-Shafer recombination rule.

Theorem 2. *Suppose the assumptions in Theorem 1 holds. We have $\tau_{i,n} = c_{i,n} + o_p(1/n)$, for each $i = 1, 2, \ldots, K$ and*

$$\left| H_n^{(DS)}(\theta, \mathbf{X}_{1,n}, \cdots, \mathbf{X}_{k,n}) - H_n^{(c)}(\theta, \mathbf{X}_{1,n}, \mathbf{X}_{2,n}, \ldots, \mathbf{X}_{k,n}) \right| \xrightarrow{P} 0. \tag{17}$$

We remark that the Assumptions 1-3 required in the two theorems are trivially satisfied for the normal example (Example 1) in Section 2.1. Arguments similar to the proof of Proposition 3 of Hannig *et al.* (2006) show that the assumptions cover the t example (Example 2) and many other examples of likelihood inference where the corresponding likelihood is asymptotically normally distributed. The assumptions do not cover examples of likelihood inference where the corresponding likelihood is not asymptotically normally distributed such as inference about parameters of a uniform distribution.

4 Numerical examples

In this section, we use both simulation and real data sets to study the properties of the combination rule based on Dempster-Shafer recombination developed in Section 2. The results are compared with corresponding CD combination methods.

4.1 Simulation studies

Example 5. This is a continuation of Example 2, the common mean problem, in Section 2.1.

We simulate k independent samples of sizes n_i from $N(\mu, \sigma_i^2)$. Without loss of generality, the common mean parameter μ is set to be $\mu = 0$. Then, based on these k independent normal samples and using (11) based on the Dempster-Shafer recombination rule, we obtain a 95% confidence interval for the common mean parameter μ. We repeat the simulation 1000 times, and computed the empirical coverage rate (the percentage of times that the 1000 confidence intervals cover the true $\mu = 0$) and the median length of the 1000 confidence intervals. The same 1000 data sets are analyzed using the corresponding normal-based and asymptotically equivalent CD combination method (3) and the no-weight rule (2).

In our simulation study, the number of studies is $k = 9$. We have considered three $\times$ two settings of parameters and sample sizes. The sample sizes $n_1 = n_2 = \ldots = n_9 \equiv n$ have three choices: I) $n = 5$, II) $n = 25$ and III) $n = 125$, representing small, medium and large sample sizes. In each of the three sets of sample sizes n, we consider two sets of variances: (a) equal variances $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_9^2 \equiv 0.01$ and (b) unequal variances $(\sigma_1^2, \ldots, \sigma_9^2) = (0.0007, 0.0007, 0.0013, 0.0009, 0.0007, 0.0004, 0.0004, 0.0002, 0.0004)$. The vari-

	DS-recombination method		Weighted CD-method		No-weight CD-method		Relative length		
Setting	Coverage	Length	Coverage	Length	Coverage	Length	CD_w vs DS	$\mathrm{CD}_{w/o}$ vs DS	$\mathrm{CD}_{w/o}$ vs CD_w
I-(a)	92%	6.36×10^{-3}	86%	5.91×10^{-3}	95%	6.19×10^{-3}	7.2%	2.7%	-4.7%
II-(a)	95%	2.68×10^{-3}	95%	2.63×10^{-3}	95%	2.65×10^{-3}	1.8%	0.9%	-0.9%
III-(a)	95%	1.17×10^{-3}	94%	1.17×10^{-3}	95%	1.17×10^{-3}	0.4%	0.2%	-0.2%
I-(b)	93%	2.80×10^{-4}	89%	2.59×10^{-4}	94%	3.12×10^{-4}	7.5%	-11.7%	-20.8%
II-(b)	95%	1.12×10^{-4}	95%	1.10×10^{-4}	94%	1.29×10^{-4}	2.1%	-14.6%	-17.1%
III-(b)	95%	4.91×10^{-5}	94%	4.85×10^{-5}	96%	5.67×10^{-5}	1.3%	-15.4%	-16.9%

Table 1: Numerical comparison of DS versus CD for the common mean problem. We report the empirical coverage, median length of 95% CIs and relative median length $(1 - a/b)$ for method-a versus method-b.

ances in (b) mimic a real data set of a key comparison by Strawderman and Rukhin (2010) studied in Section 4.2. The numerical results are reported in Table 1.

Table 1 suggests that the approach applying the Dempster-Shafer recombination rule and the norm-based weighted CD combination approach (3) have almost the identical performance when sample sizes are large. The weighted CD combination approach (3) typically provides shorter confidence intervals, but when the sample sizes are small the approach has an under-coverage problem. The approach using the Dempster-Shafer recombination rule also has a slight under-coverage problem when sample sizes are small, but their performance are better than the weighted CD combination approach (3). The norm-based no-weight CD combination approach (2) can produce intervals at right coverage at all cases. But when samples have heterogeneous variances, the no-weight CD-combination approach (2) produces longer confidence intervals, suggesting a loss of efficiency. This loss of efficiency will not be diminish even when sample sizes go to ∞.

Example 6. This is a continuation of Example 4, common odds ratio, in Section 2.2.

We simulate k sets of independent sample (x_{1i}, x_{2i}) from two independent Binomial distributions $x_{1i} \sim Binomial(n_{1,i}, p_{1i})$ and $x_{0i} \sim Binomial(n_{0,i}, p_{0i})$, for $i = 1, \ldots, k$. To ensure that we have the common odds ratio, say ψ, across all k studies, we set values for ψ and p_{1i}, and compute p_{0i} by $p_{0i} = e^{\pi}_{0i}/(1+e^{\pi}_{0i})$ with $\pi_{0i} = \{p_{1i}/(1-p_{1i})\}/\psi$, for $i = 1, \ldots, k$. Then, based on these k sets of paired binomial samples and using (15) based on the Dempster-Shafer recombination rule, we obtain a 95% confidence interval for the log common odds ratio $\log(\psi)$. We repeat the simulation 1000 times, and compute the empirical coverage rate and the median length of the 1000 confidence intervals. The same 1000 data sets are analyzed using the asymptotically equivalent

CD combination method (3).

In this simulation example, the number of studies is $k = 6$. We have considered three $\times$ five = 15 different settings of parameters and sample sizes. Five values of the true common odds ratio ψ are considered: $\psi = 0, 3, 6, 1/3, 1/6$. For each of the five ψ values, we have considered thee sample sizes and p_{1i} settings: I) $n_1 = \ldots = n_6 = 20$, $m_1 = \ldots = m_6 = 20$, $(p_{11}, p_{12}, p_{13}, p_{14}, p_{15}, p_{16}) = (0.10, 0.20, 0.10, 0.05, 0.10, 0.15)$; II) $(n_1, n_2, n_3, n_4, n_5, n_6) = (39, 44, 107, 103, 110, 154)$, $(m_1, m_2, m_3, m_4, m_5, m_6) = (43, 44, 110, 100, 106, 146)$, $(p_{11}, p_{12}, p_{13}, p_{14}, p_{15}, p_{16}) = (0.0513, 0.0909, 0.0561, 0.0680, 0.0636, 0.0714)$; III) $n_1 = n_2 = \ldots = n_6 = 200$, $m_1 = m_2 = \ldots = m_6 = 200$, $(p_{11}, p_{12}, p_{13}, p_{14}, p_{15}, p_{16}) = (0.10, 0.20, 0.10, 0.05, 0.10, 0.15)$. The numerical results are reported in Table 2.

Table 2 suggests that the approach applying the Dempster-Shafer recombination rule and the norm-based weighted CD combination approach (3) provide more or less very similar results, in terms of the coverage rate and interval length. The results are getting closer and closer when the sample sizes increases. Unlike the previous example, the approach based on DS recombination rule produces slightly shorter intervals in majority settings, although the computing based on DS recombination rule is much more complicated in the discrete settings. The equivalent CD recombined method applies the half correction to each of the individual confidence distribution before combining them while the Dempster-Shafer rule first recombines the confidence random sets and then applies the half correction to the final results.

4.2 Real data examples

We end with analysis of two real data sets. One concerns pp′-DDT levels in fish-oil measured by nine laboratories (Webb *et al.*, 2003; Strawderman and Rukhin, 2010) and the other is an analysis of mortality data for control and intravenous lidocaine treatment from six studies (Normand, 1999).

For increased clarity, we illustrate our numerical results by plotting confidence curves (Birnbaum, 1961). For a given confidence distribution $H(\theta, \mathbf{x})$, its corresponding confidence curve is defined as

$$CV(\theta) = 1 - 2|H(\theta, \mathbf{x}) - 0.5| = 2 \min\Big\{H(\theta, \mathbf{x}), 1 - H(\theta, \mathbf{x})\Big\}.$$

On a plot of $CV(\theta)$ versus θ, a line across the height (y-axis) of α, for any $0 < \alpha < 1$, intersects with the confidence curve at two points, and these

Setting	Odds Ratio	DS-recombination method Coverage	Length	Weighted CD-method Coverage	Length	Relative length CD vs DS
I	1	96%	1.66	98%	1.71	3.3%
II	1	95%	0.959	97%	0.981	2.3%
III	1	95%	0.505	95%	0.509	0.7%
I	3	96%	2.24	97%	2.12	-5.3%
II	3	95%	1.33	97%	1.35	1.4%
III	3	96%	0.669	97%	0.676	1.1%
I	6	98%	3.23	91%	2.35	-27.2%
II	6	94%	1.78	93%	1.68	-5.5%
III	6	96%	0.861	95%	0.868	0.8%
I	1/3	96%	1.44	97%	1.47	2.5%
II	1/3	95%	0.802	96%	0.813	1.3%
III	1/3	95%	0.442	95%	0.444	0.5%
I	1/6	95%	1.40	97%	1.42	1.7%
II	1/6	96%	0.765	95%	0.772	0.9%
III	1/6	94%	0.432	95%	0.433	0.4%

Table 2: Numerical comparison of DS versus CD for the 2×2 table setting. We report the empirical coverage, median length of 95% CIs and the relative median length $(1 - a/b)$ for method-a (CD) versus method-b (DS).

two points correspond (on x-axis) to a $1-\alpha$ level, equal tailed, two sided confidence interval for θ. Thus, a confidence curve is a graphical device that shows confidence intervals of all levels; see, e.g. Birnbaum (1961); Bender *et al.* (2005). The mode of a confidence curve plot $\hat{\theta} = \arg\max_\theta CV(\theta) = H^{-1}(1/2)$ is the median of the confidence distribution. It provides a point estimator which is typically a median unbiased (Birnbaum, 1961) and consistent under some mild condition (Singh *et al.*, 2007; Xie and Singh, 2012).

Example 7. An interlaboratory study CCQM-K21 involving nine national laboratories across nine different countries (Webb *et al.*, 2003) reported concentrations of pesticide pp′-DDT in fish-oil collected by the nine national laboratories, with each making replicate measurements on aliquots of fish-oil. Table 3 is a reproduction of the nine means and standard errors reported in Table 1a of Webb *et al.* (2003), along with the reported sample sizes n_i. A consensus (or reference) value is required to be established by combining information from the results of these nine laboratories. Strawderman and Rukhin (2010) studied the point estimation problem for the data. We provide here combined confidence distributions (distributional estimation) using the the Dempster-Shafer recombination rule (11), the asymptotically equivalent CD combination rule (3), and also the CD combination rule without

Sample size (n_i)	4	3	4	5	4	4	4	4	4
Mean ($\mu g g^{-1}$)	.0732	.0794	.0756	.0736	.0711	.0739	.0725	.0724	.0768
SE ($\mu g g^{-1}$)	.0007	.0007	.0013	.0009	.0007	.0004	.0004	.0002	.0004

Table 3: CCQM-K21 data on pp′-DDT in fish-oil for nine laboratories.

any weight (2).

Figure 1 plots the confidence curves obtained from the Dempster-Shafer recombination rule, the two CD combination methods and the individual confidence curve from data collected in each laboratory. The plots indicate that all combination rules provide a good aggregation of the information for the nine individual laboratories. The combined confidence curves by the three methods appear close, although the recombined confidence curve by the Dempster-Shafer recombination rule is slightly skewed and also slighted shifted to the left and the combined confidence curve by the no-weight CD combination (2) is slightly wider and also slightly shifted to the right. The point estimate of the common mean by the three methods are 0.0727, 0.0732 and 0.0736, with corresponding 95% confidence intervals $(0.0723, 0.0736)$, $(0.0726, 0.0740)$ and $(0.0728, 0.0745)$, respectively.

Example 8. Table 1 of Normand (1999) contained mortality data for control and intravenous lidocaine treatment from $k = 6$ studies. The sample sizes of these six studies range from 82 to 300 heart attack patients. A parameter of interest is the logarithm of the odds ratio parameter of the treatment versus control. In this example, we obtain and compare the combined estimators of the common odds ratio using both the combination rule based on Dempster-Shafer recombination (15) and the normal based CD combination rule (3).

Figure 2 below plots the confidence curves obtained from the Dempster-Shafer recombination rule (15) and its asymptotically equivalent CD combination (3), as well as the individual confidence curve from data collected in each clinical trial. For each individual trial, the confidence distribution used is the p-value function from the left-sided test using the Fisher exact test but with half correction, as stated in Section 2.2. Based on Figure 2, we see again that both combined inferences provide a good aggregation of of the information for the six individual clinical trials. Again, the combined confidence curves by the two different methods are very similar. The point estimate of the common log odds ratio by the two methods are 0.575 and 0.568,

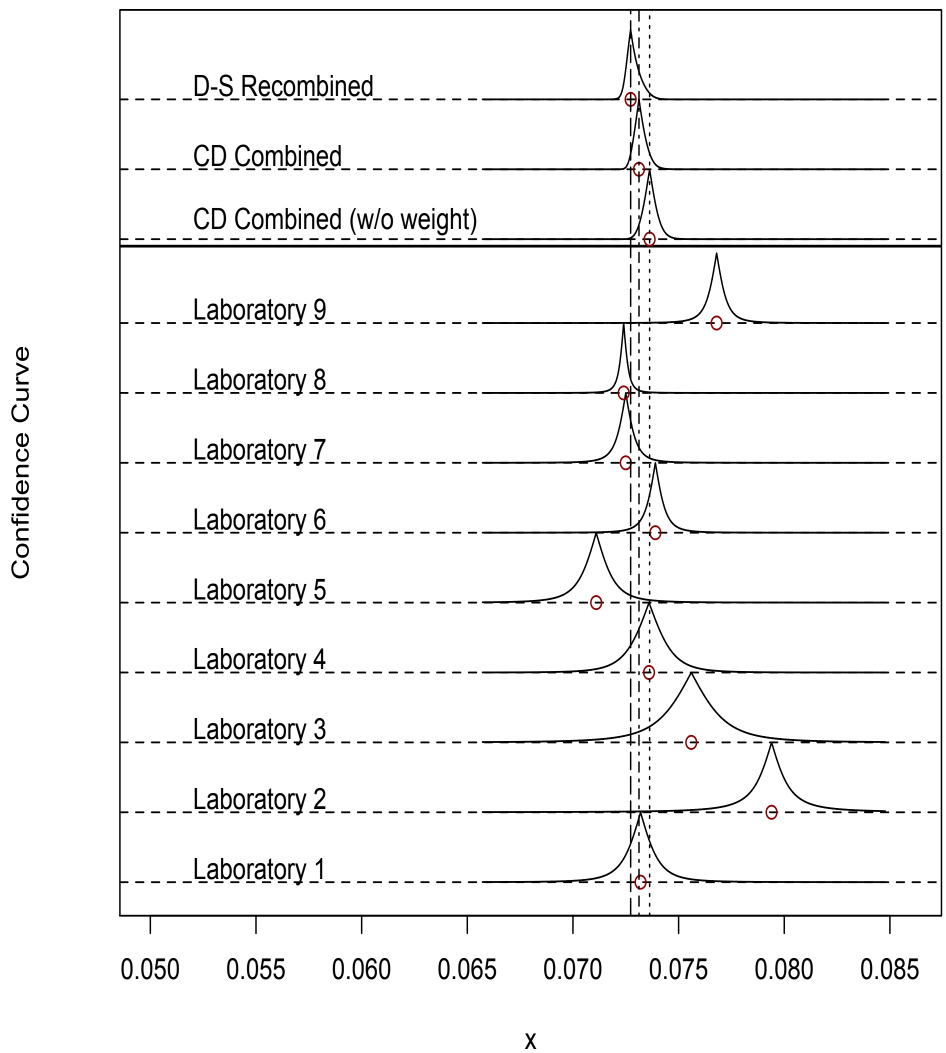

Figure 1: The comparison of confidence curves for combination of several confidence distributions for the common normal mean for data in Example 7. The plot displays confidence curves combined using the Dempster-Shafer based rule (11), the asymptotically equivalent normal based CD combination (3), the CD combination rule without weight (2) and the individual confidence curves that are being combined. The red circles denote the median of each of the confidence distributions.

with corresponding 95% confidence intervals $(0.010, 1.126)$ and $(0.033, 1.142)$, respectively.

Acknowledgments

The authors thank the editor, AE and two referees for their constructive comments and suggestions, which helped improve the quality of the paper. The authors also thank Guang Yang and Pixu Shi for their assistance in computer programming.

Appendix: Proofs

Proof of Theorem 1. Assumption 1 and some algebra imply that ${c_{i,n}}^{-1}\{t_{i,n}(\mathbf{X}_{i,n}) - \theta_0\} \xrightarrow{\mathcal{D}} N(0,1)$. Consequently $(\sum_{j=1}^k c_{j,n}^{-2})^{1/2}(T_n^{(DS)} - \theta_0) \xrightarrow{\mathcal{D}} N(0,1)$.

We will now investigate the right-hand-side of (10). Define $\mathbf{X}_n = (\mathbf{X}_{1,n}, \ldots, \mathbf{X}_{k,n})$ and

$$K_n(\mathbf{X}_n) = (2\pi)^{-\frac{k-1}{2}} \exp\left(\sum_{i=1}^k \frac{\left\{ t_{i,n}(\mathbf{X}_{i,n}) - T_n^{(DS)} \right\}^2}{2c_{i,n}^2} \right).$$

Notice that $K_n(\mathbf{X}_n)$ is bounded in probability and

$$\tilde{h}_n(\theta | T_n^{(DS)}) = K_n(\mathbf{X}_n) \left(\sum_{i=1}^k c_{i,n}^{-2} \right)^{1/2} \prod_{i=1}^n \varphi\left(\frac{\theta - t_{i,n}(\mathbf{X}_{i,n})}{c_{i,n}} \right).$$

Fix $j = 1, \ldots, k$ and denote the term

$$\tilde{h}_{j,n}^{(DS)} = K_n(\mathbf{X}_n) \left(\sum_{i=1}^k c_{i,n}^{-2} \right)^{1/2} c_{j,n} \left| \frac{\partial}{\partial\theta} H_{j,n}(\theta, \mathbf{X}_{j,n}) \right| \prod_{\substack{i=1 \\ i \neq j}}^k \left| c_{i,n} D_{\mathbf{x}_{i,n}} H_{i,n}(\theta, \mathbf{X}_{i,n}) \right|.$$

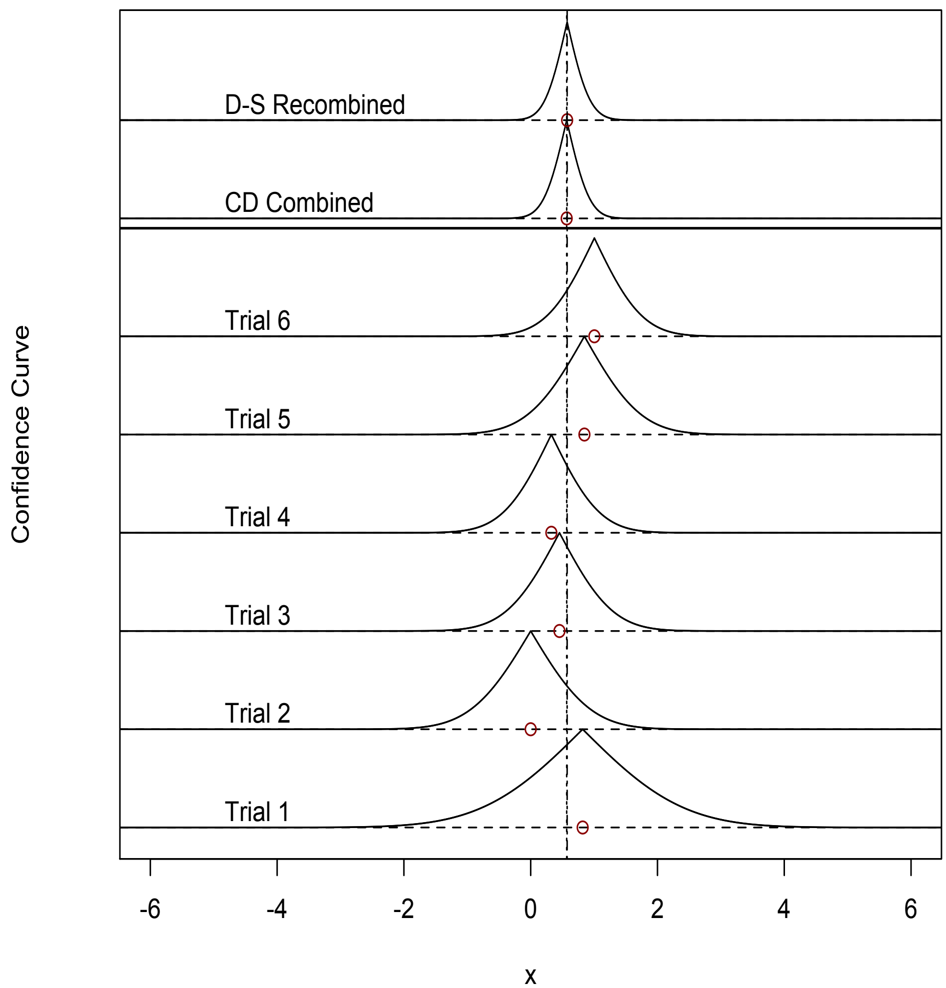

Figure 2: The comparison of confidence curves for combination of several confidence distributions for the logarithm of the odds ratio for data in Example 8. The plot displays confidence curves combined using the Dempster-Shafer based rule (15), the normal based CD combination rule (3) and the individual confidence curves that are being combined. The red circles denote the median of each of the confidence distributions.

Compute

$$\left|\tilde{h}^{(DS)}_{j,n} - \tilde{h}_n(\theta|T^{(DS)}_n)\right|$$

$$\leq \prod_{\substack{i=1\\i\neq j}}^{k} \left|c_{i,n} D_{\mathbf{x}_{i,n}} H_{i,n}(\theta, \mathbf{X}_{i,n})\right|$$

$$\times K_n(\mathbf{X}_n) c_{j,n} \left(\sum_{j=1}^{k} c_{j,n}^{-2}\right)^{1/2} \left|\frac{\partial}{\partial\theta} H_{j,n}(\theta, \mathbf{X}_{j,n}) - \frac{1}{c_{j,n}} \varphi\left(\frac{\theta - t_{j,n}(\mathbf{X}_{j,n})}{c_{j,n}}\right)\right|$$

$$+ \prod_{\substack{i=1\\i\neq j}}^{k-1} \left|c_{i,n} D_{\mathbf{x}_{i,n}} H_{i,n}(\theta, \mathbf{X}_{i,n})\right| \varphi\left(\frac{\theta - t_{j,n}(\mathbf{X}_{j,n})}{c_{j,n}}\right)$$

$$\times K_n(\mathbf{X}_n) c_{k,n} \left(\sum_{j=1}^{k} c_{j,n}^{-2}\right)^{1/2} \left|\frac{\partial}{\partial\theta} H_{k,n}(\theta, \mathbf{X}_{k,n}) - \frac{1}{c_{k,n}} \varphi\left(\frac{\theta - t_{k,n}(\mathbf{X}_{k,n})}{c_{k,n}}\right)\right|$$

$$+ \prod_{\substack{i=1\\i\neq j}}^{k-2} \left|c_{i,n} D_{\mathbf{x}_{i,n}} H_{i,n}(\theta, \mathbf{X}_{i,n})\right| \varphi\left(\frac{\theta - t_{k,n}(\mathbf{X}_{k,n})}{c_{k,n}}\right) \varphi\left(\frac{\theta - t_{j,n}(\mathbf{X}_{j,n})}{c_{j,n}}\right)$$

$$\times K_n(\mathbf{X}_n) c_{k-1,n} \left(\sum_{j=1}^{k} c_{j,n}^{-2}\right)^{1/2} \left|\frac{\partial}{\partial\theta} H_{k-1,n}(\theta, \mathbf{X}_{k-1,n}) - \frac{1}{c_{k-1,n}} \varphi\left(\frac{\theta - t_{k-1,n}(\mathbf{X}_{k-1,n})}{c_{k-1,n}}\right)\right|$$

$$+ \dots$$

integral of which is converging to 0 by Assumption 2. Equation (16) follows by uniform integrability arguments (Durrett, 2005, Theorem 5.2) and summing over j.

To prove that the Dempster-Shafer recombined confidence distribution is an asymptotic confidence distribution notice that

$$\int_{-\infty}^{\theta_0} h_n^{(DS)}(\theta|\mathbf{X}_{1,n}, \cdots, \mathbf{X}_{k,n})\, d\theta = \int_{-\infty}^{\theta_0} \tilde{h}_n(\theta|T_n^{(DS)})\, d\theta + \varepsilon_n,$$

where

$$|\varepsilon_n| \leq \int_\Theta \left|h_n^{(DS)}(\theta|\mathbf{X}_{1,n}, \cdots, \mathbf{X}_{k,n}) - \tilde{h}_n(\theta|T_n^{(DS)})\right| d\theta \xrightarrow{P} 0$$

and

$$\int_{-\infty}^{\theta_0} \tilde{h}_n(\theta|T_n^{(DS)})\, d\theta = \Phi\{(\sum_{j=1}^{k} c_{j,n}^{-2})^{1/2}(\theta_0 - T_n^{(DS)})\} \xrightarrow{\mathcal{D}} U(0,1),$$

since $(\sum_{j=1}^{k} c_{j,n}^{-2})^{1/2}(T_n^{(DS)} - \theta_0) \xrightarrow{\mathcal{D}} N(0,1)$. The statement now follows. □

Proof of Theorem 2. For a given β, $0 < \beta < 1$, by basic calculations and Assumption 2b, we have

$$\begin{aligned}
&\Big|H_{i,n}\Big(t_{j,n}(\mathbf{X}_{i,n}) + c_{i,n}\Phi^{-1}(\beta)\Big) - \beta\Big| \\
&= \Big| \int_{-\infty}^{t_{i,n}(\mathbf{X}_{i,n})+c_{i,n}\Phi^{-1}(\beta)} dH_{i,n}(\theta, \mathbf{X}_{i,n}) - \int_{-\infty}^{t_{i,n}(\mathbf{X}_{i,n})+c_{i,n}\Phi^{-1}(\beta)} \frac{1}{c_{i,n}}\varphi\left(\frac{\theta - t_{i,n}(\mathbf{X}_{i,n})}{c_{i,n}}\right) d\theta\Big| \\
&\le \int_{-\infty}^{t_{i,n}(\mathbf{X}_{i,n})+c_{i,n}\Phi^{-1}(\beta)} \Big|\frac{\partial}{\partial\theta}H_{i,n}(\theta, \mathbf{X}_{i,n}) - \frac{1}{c_{i,n}}\varphi\left(\frac{\theta - t_{i,n}(\mathbf{X}_{i,n})}{c_{i,n}}\right)\Big| d\theta \\
&\le \int_{\Theta} \Big|\frac{\partial}{\partial\theta}H_{i,n}(\theta, \mathbf{X}_{i,n}) - \frac{1}{c_{i,n}}\varphi\left(\frac{\theta - t_{i,n}(\mathbf{X}_{i,n})}{c_{i,n}}\right)\Big| d\theta \xrightarrow{P} 0.
\end{aligned}$$

Since we assume $H_i(\theta, \mathbf{X}_{i,n})$ is continuous in θ, we have $H_i^{-1}(\beta) = t_{i,n}(\mathbf{X}_{i,n}) + c_{i,n}\Phi^{-1}(\beta) + o_p(1)$. Thus, $\tau_{i,n} = \{H_{i,n}^{-1}(.75) - H_{i,n}^{-1}(.25)\}/\{2\Phi^{-1}(.75)\} = c_{i,n} + o_p(1)$. The first statement of the theorem follows.

Now, write $\epsilon_{i,n} = H_{i,n}(\theta_0, \mathbf{X}_{i,n}) - \Phi[\{\theta_0 - t_{i,n}(\mathbf{X}_{i,n})\}/c_{i,n}]$. By Assumption 1b, $\epsilon_{i,n} \xrightarrow{P} 0$. From Lemma 1 of Xie *et al.* (2011) and also noting that $\tau_{i,n} = c_{i,n} + o_p(1)$, we have

$$\begin{aligned}
&\Big|H_n^{(c)}\big(\theta, \mathbf{X}_{1,n}, \mathbf{X}_{2,n}, \ldots, \mathbf{X}_{k,n}\big) - \Phi\{(\sum_{j=1}^{k} c_{j,n}^{-2})^{1/2}(T_n^{(DS)} - \theta_0)\}\Big| \\
&= \Big|\Phi\left(\frac{\sum_{i=1}^{k} \tau_{i,n}^{-1}\Phi^{-1}\{H_i(\theta, \mathbf{X}_{1,n})\}}{(\sum_{i=1}^{k} \tau_{i,n}^{-2})^{1/2}}\right) - \Phi\left(\frac{\sum_{i=1}^{k} c_{i,n}^{-1}\Phi^{-1}\{\Phi(\{\theta - t_{i,n}(\mathbf{X}_{i,n})\}/c_{i,n})\}}{(\sum_{i=1}^{k} c_{i,n}^{-2})^{1/2}}\right)\Big| \\
&\le \sum_{i=1}^{k} \epsilon_{i,n} + o_p(1) \xrightarrow{P} 0.
\end{aligned}$$

Thus, $H_n^{(c)}\big(\theta, \mathbf{X}_{1,n}, \mathbf{X}_{2,n}, \ldots, \mathbf{X}_{k,n}\big)$ and

$$\Phi\{(\sum_{j=1}^{k} c_{j,n}^{-2})^{1/2}(\theta - T_n^{(DS)})\} = \int_{-\infty}^{\theta} \tilde{h}_n(\eta|T_n^{(DS)})\, d\eta$$

are asymptotically equivalent. The second statement of the theorem follows from equation (16). □

References

Bender, R., Berg, G. and Zeeb, H. (2005) Tutorial: Using confidence curves in medical research. *Biometrical Journal*, **47**, 237–247.

Birnbaum, A. (1961) Confidence curves: An omnibus technique for estimation and testing statistical hypotheses. *J. Amer. Statist. Assoc.*, **56**, 246–249.

Casella, G. and Berger, R. L. (2002) *Statistical Inference*. Pacific Grove, CA: Wadsworth and Brooks/Cole Advanced Books and Software, 2nd edn.

Cox, D. (1958) Some problems with statistical inference. *The Annals of Mathematical Statistics*, **29**, 357–372.

Dempster, A. P. (2008) The Dempster-Shafer Calculus for Statisticians. *International Journal of Approximate Reasoning*, **48**, 365–377.

Durrett, R. (2005) *Probability: theory and examples*. Duxburry Advanced Series. Brooks/Cole, third edn.

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

— (1998) R. A. Fisher in the 21st century. *Statist. Sci.*, **13**, 95–122. With comments and a rejoinder by the author.

Fisher, R. (1973) *Statistical Methods and Scientific Inference (3rd edition)*. New York: Hafner Press.

Hannig, J. (2009) On Generalized Fiducial Inference. *Statistica Sinica*, **19**, 491–544.

— (2012) Generalized Fiducial Inference via Discretizion. *Statistica Sinica*.

Hannig, J., Iyer, H. K. and Patterson, P. (2006) Fiducial generalized confidence intervals. *Journal of American Statistical Association*, **101**, 254 – 269.

Liu, D., Liu, R. and Xie, M. (2011) Exact meta-analysis approach for the common odds ratio of 2 by 2 tables with rare events. *Technical Report, Department of Statistics, Rutgers University.*

Normand, S.-L. (1999) Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine*, **18**, 321–359.

Schweder, T. and Hjort, N. L. (2002) Confidence and likelihood. *Scand. J. Statist.*, **29**, 309–332. Large structured models in applied sciences; challenges for statistics (Grimstad, 2000).

Shafer, G. (1976) *A mathematical theory of evidence*. Princeton, New Jersey: Princeton University Press.

Singh, K., Xie, M. and Strawderman, W. E. (2001) Confidence distributions – concept, theory and applications. *Tech. rep.*, Department of Statistics, Rutgers University. Updated 2004.

— (2005) Combining information from independent sources through confidence distributions. *The Annals of Statistics*, **33**, 159–183.

— (2007) Confidence distribution (cd)-distribution estimator of a parameter. *IMS Lecture Notes-Monograph Series*, **54**, 132–150.

Strawderman, W. and Rukhin, A. (2010) Simultaneous estimation and reduction of nonconformity in inter-laboratory studies. *J. R. Statist. Soc. B*, **72**, 219–234.

Tian, L., Cai, T., Pfeffer, M. A., Piankov, N., Cremieux, P. Y. and Wei, L. J. (2009) Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent $2\times$ 2 tables with all available data but without artificial continuity correction. *Biostatistics*, **10**, 275–281.

Wang, J. C.-M., Hannig, J. and Iyer, H. K. a. (2012) Fiducial prediction intervals. *Journal of Statistical Planning and Inference*, **142**, 1980 – 1990.

Webb, K. S., Carter, D. and Wolff Briche, C. S. J. (2003) Ccqm-k21: Key comparison of the determination of pp′-ddt in fish oil, final report. *Metrologia*, **40**, tech. suppl. 08004.

Xie, M. and Singh, K. (2012) Confidence distribution, the frequentist distribution estimator of a parameter – a review. *International Statistical Reviews. To appear (Invited review article with discussions).*

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

Zhang, J. and Liu, C. (2011) Dempster-Shafer inference with weak beliefs. *Statistica Sinica*, **21**, 475–494.