International Statistical Review
Revue Internationale de Statistique

International Statistical Review (2013), 81, 1, 3–39 doi:10.1111/insr.12000

Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review

Min-ge Xie and Kesar Singh

Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854, USA
E-mail: mxie@stat.rutgers.edu

Summary

In frequentist inference, we commonly use a single point (point estimator) or an interval (confidence interval/"interval estimator") to estimate a parameter of interest. A very simple question is: *Can we also use a distribution function ("distribution estimator") to estimate a parameter of interest in frequentist inference in the style of a Bayesian posterior?* The answer is affirmative, and confidence distribution is a natural choice of such a "distribution estimator". The concept of a confidence distribution has a long history, and its interpretation has long been fused with fiducial inference. Historically, it has been misconstrued as a fiducial concept, and has not been fully developed in the frequentist framework. In recent years, confidence distribution has attracted a surge of renewed attention, and several developments have highlighted its promising potential as an effective inferential tool.

This article reviews recent developments of confidence distributions, along with a modern definition and interpretation of the concept. It includes distributional inference based on confidence distributions and its extensions, optimality issues and their applications. Based on the new developments, the concept of a confidence distribution subsumes and unifies a wide range of examples, from regular parametric (fiducial distribution) examples to bootstrap distributions, significance (*p*-value) functions, normalized likelihood functions, and, in some cases, Bayesian priors and posteriors. The discussion is entirely within the school of frequentist inference, with emphasis on applications providing useful statistical inference tools for problems where frequentist methods with good properties were previously unavailable or could not be easily obtained. Although it also draws attention to some of the differences and similarities among frequentist, fiducial and Bayesian approaches, the review is not intended to re-open the philosophical debate that has lasted more than two hundred years. On the contrary, it is hoped that the article will help bridge the gaps between these different statistical procedures.

Key words: Confidence distribution; statistical inference; fiducial distribution; Bayesian method; likelihood function; estimation theory.

1 Introduction

In Bayesian inference, researchers typically rely on a posterior distribution to make inference on a parameter of interest, where the posterior is often viewed as a "distribution estimator" for the parameter. A nice aspect of using a distribution estimator is that it contains a wealth of information for almost all types of inference. In frequentist inference, however, we often use a single point (point estimator) or an interval (confidence interval/"interval estimator") to estimate a parameter of interest. A simple question is:

Can we also use a distribution function, or a "distribution estimator", to estimate a parameter of interest in frequentist inference in the style of a Bayesian posterior?

The answer is affirmative and, in fact, *confidence distribution* (CD) is one such a "distribution estimator", that can be defined and interpreted in a frequestist (repetition) framework, in which the parameter is a fixed and non-random quantity.

The concept of confidence distribution has a long history, especially with its early interpretation associated with fiducial reasoning (see, e.g., Fisher, 1973; Efron, 1993; Cox, 2006; Hampel, 2006). Historically, it has been long misconstrued as a fiducial concept, and has not been fully developed under the frequentist framework—perhaps partly due to Fisher's "stubborn insistence" and his "unproductive dispute" with Neyman (Zabell, 1992). In recent years, the confidence distribution concept has attracted a surge of renewed attention, and the recent developments have based on a redefinition of the confidence distribution as a *purely frequentist concept, without any fiducial reasoning*. The goal of these new developments is not to derive a new fiducial theory that is paradox free. Rather, it is on providing a useful statistical tool for problems where frequentist methods with good properties were previously unavailable or could not be easily obtained. One nice aspect of treating a confidence distribution as a purely frequentist concept is that the confidence distribution is now a clean and coherent frequentist concept (similar to a point estimator) and it frees itself from those restrictive, if not controversial, constraints set forth by Fisher on fiducial distributions.

A confidence distribution can often be loosely referred to as a sample-dependent distribution that can represent confidence intervals of all levels for a parameter of interest. One such distribution estimator, that is well known and extremely popular in modern statistical application, is Efron's bootstrap distribution, albeit the concept of a confidence distribution is much broader. Efron (1998) stated that a bootstrap distribution is a "distribution estimator" and a "confidence distribution" of the parameter that it targets. Clearly, the implementation of a bootstrap method is done entirely within the frequentist domain and it does not involve any fiducial or Bayesian reasoning. The same is true for a confidence distribution.

A basic example, which was also used by Fisher (1935, 1973) to illustrate his fiducial function, is from the normal mean inference problem with sample $x_i \sim N(\mu, \sigma^2), i = 1, \ldots, n$. Under this setting, $N(\bar{x}, \sigma^2)$ or more formally in its cumulative distribution function form $H_\Phi(\mu) = \Phi(\sqrt{n}(\mu - \bar{x})/\sigma)$ is a "distribution estimator" or a confidence distribution for μ, when σ^2 is known. See Example 1 later for more details. Also, the distribution function $H_t(\mu) = F_{t_{n-1}}(\sqrt{n}(\mu - \bar{x})/s)$ can be used to estimate and make inference for μ, when σ^2 is not known, and $N(\bar{x}, s^2)$ can be used to estimate and make inference for μ when the sample size n is large, regardless of whether σ^2 is known or not. Here, $\bar{x}$ and s^2 are the sample mean and variance, and Φ and $F_{t_{n-1}}$ stand for the cumulative distribution functions of the standard normal and the t-distribution with $n - 1$ degrees of freedom, respectively.

The introduction of the confidence distribution concept is arranged in two sections. Section 2 reviews and introduces the concept of confidence distribution, along with some examples. Section 3 provides further in-depth discussions on the concept, in which we underscore an intimate connection between bootstrap and confidence distributions, and explore the relations and distinctions between the confidence distributions, fiducial distributions and Bayesian inference. Specifically, Section 2.1 briefly reviews the history of confidence distribution and its classical definition. Section 2.2 provides a modern definition of a confidence distribution, and

Section 2.3 provides several illustrative examples. Section 3.1 introduces the concept of *CD-random variable* and explores an underlying similarity between inference based on a general confidence distribution and inference based on a bootstrap distribution. Section 3.2 notes that confidence and fiducial distributions have been always linked since their inception, and provides a further discussion which brings out the intertangled relationship between confidence and fiducial distributions. Section 3.3 is devoted to stress a unity between confidence distribution and Bayesian inference and also their differences. From Sections 2 and 3, we can see that the concept of confidence distribution is purely a frequenstist notion. It is also very broad and subsumes many well-known notions in statistics. Indeed, a main theme throughout the paper is that any approach that can build confidence intervals for all levels, regardless of whether they are exact or asymptotically justified, can potentially be unified under the confidence distribution framework. The approaches discussed in Sections 2 and 3 include the classical fiducial examples from pivotal functions, p-value functions from one-sided hypothesis tests, bootstrap distributions, likelihood approaches and, even sometimes, the Bayesian method.

A confidence distribution is a probability distribution function on the parameter space. It contains a wealth of information for inference; much more than a point estimator or a confidence interval. For instance, in the above normal example, the mean/median/mode of the distribution estimator $N(\bar{x}, \sigma^2/n)$ when σ^2 is known provides a point estimator $\bar{x}$ for μ, and its $\alpha/2$ and $1-\alpha/2$ quantiles $\big(H_\Phi^{-1}(\alpha/2), H_\Phi^{-1}(1-\alpha/2)\big) = \big(\bar{x} + \Phi^{-1}(\alpha/2)\sigma/\sqrt{n},\ \bar{x} + \Phi^{-1}(1-\alpha/2)\sigma/\sqrt{n}\big)$ provide a level $(1-\alpha)100\%$ confidence interval, for every $0 < \alpha \leq 1$. Furthermore, its tail mass $H_\Phi(b) = \Phi\big\{(b-\bar{x})/(\sigma/\sqrt{n})\big\}$ provides a p-value for the one-sided hypothesis test $K_0 : \mu \leq b$ versus $K_1 : \mu > b$, for a given b. In Section 4, we provide a systematic review on how we obtain various types of inference under the frequentist framework from a confidence distribution. It reveals that the CD-based inference is similar in style to that of a Bayesian posterior. It also underscores the well-known duality between tests and confidence sets in which one can be derived from the other and vice-versa.

Unlike classical fiducial inference, more than one confidence distributions may be available to estimate a parameter under any specific setting ("non-uniqueness" in estimation). A natural question related to decision theory is "which one is better?" or "which one to use?" Section 5 reviews some optimality results for confidence distributions. These findings suggest that, in many cases, there are optimality results for confidence distributions just like those in point estimation theory. Depending on the setting and the criterion used, sometimes there is an unique "best" (in terms of optimality) confidence distribution and sometimes there is no optimal confidence distribution available. This is not different from the practice of point estimation, but is certainly different from the practice of fiducial inference.

The concept of a confidence distribution has a long history. But perhaps due to its historic connection to fiducial distribution, little attention has been paid to it in the past, especially in applications. This has changed in recent years with many new developments emphasizing applications of CD-based inferences. In Sections 6 and 7, we provide a review of several recent studies involving confidence distributions, ranging from combination of information from independent studies, incorporation of expert opinions with data in a frequentist setting, approximate likelihood calculations, confidence curves, CD-based Monte Carlo methods to applications in financial statistics, biomedical research and others. Section 6 focuses on the topic of combining information from independent sources, with which our group is most involved. Section 7 reviews several developments on CD-based likelihood calculations, confidence curves and CD-based Monte Carlo methods, and also provides a survey of several more-applied applications. From the examples in Sections 6 and 7, we can see that the CD-based inference can provide useful statistical approaches for many problems and it has broad applications.

To end this Introduction section, we cite from Efron (1998) on Fisher's contribution of the fiducial distribution which is quite relevant in the context of confidence distributions: "... but here is a safe prediction for the 21st century: statisticians will be asked to solve bigger and more complicated problems. I believe there is a good chance that objective Bayes methods will be developed for such problems, and that something like fiducial inference will play an important role in this development. Maybe Fisher's biggest blunder will become a big hit in the 21st century!" It is our hope that recent emerging developments on confidence distributions, along with recent surge of publications on generalized fiducial distributions and objective Bayes, will stimulate further explorations that can enrich statistical sciences.

2 The Concept of Confidence Distribution

2.1 A Classical Definition and the History of the CD Concept

The concept of "confidence" was first introduced by Neyman (1934, 1937) in his seminal papers on confidence intervals, where frequentist repetition properties for confidence were clarified. According to Fraser (2011), the seed idea of a confidence distribution can be even traced back before Neyman (1934, 1937) to Bayes (1763) and Fisher (1922). The earliest use of the terminology "confidence distribution" that we can find so far in a formal publication is Cox (1958). But the terminology appears to have been used before, as early as in 1927; c.f., David & Edwards (2000), p. 191, for an excerpt of a letter from E.S. Pearson to W.S. Gossett. Schweder & Hjort (2002) and Hampel (2006) suggested that confidence distributions are "the Neymanian interpretation of Fisher's fiducial distributions", although Fisher furiously disputed this interpretation.

As in the case of a fiducial distribution, for which Fisher did not give a general definition, we can not find a precise yet general definition for a confidence distribution in the classical literature. But, when discussing interval estimation, Cox (1958) suggested that a confidence distribution "can either be defined directly, or can be introduced in terms of the set of all confidence intervals at different levels of probability". Cox (1958) further stated that "Statements made on the basis of this distribution, provided we are careful about their form, have a direct frequency interpretation". The most commonly used approach to describe the concept of a confidence distribution in the literature is via inverting the upper limits of a whole set of lower side confidence intervals, often using some special examples. The following two paragraphs are from Efron (1993), where he defined a confidence distribution using a set of asymptotic normal confidence intervals. Similar approaches can be found in Cox (2006), among others.

> Suppose that a data set x is observed from a parametric family of densities $g_\mu(x)$, depending on an unknown parameter vector μ, and that inferences are desired for $\theta = t(\mu)$, a real-valued function of μ. Let $\theta_x(\alpha)$ be the upper endpoint of an exact or approximate one-sided level-α confidence interval for θ. The standard intervals for example have
>
> $$\theta_x(\alpha) = \hat{\theta} + \hat{\sigma} z^{(\alpha)},$$
>
> where $\hat{\theta}$ is the maximum likelihood estimate of θ, $\hat{\sigma}$ is the Fisher information estimate of standard error for $\hat{\theta}$, and $z^{(\alpha)}$ is the α-quantile of a standard normal distribution, $z^{(\alpha)} = \Phi^{-1}(\alpha)$. We write the inverse function of $\theta_x(\alpha)$ as $\alpha_x(\theta)$, meaning the value of α corresponding to upper endpoint θ for the confidence interval, and assume that $\alpha_x(\theta)$ is smoothly increasing

in θ. For the standard intervals, $\alpha_x(\theta) = \Phi((\theta - \hat{\theta})/\hat{\sigma})$, where Φ is the standard normal cumulative distribution function.

The confidence distribution for θ is defined to be the distribution having density

$$\pi_x^{\dagger}(\theta) = d\alpha_x(\theta)/d\theta. \qquad (1.2)_e$$

We shall call $(1.2)_e$ the confidence density. This distribution assigns probability 0.05 to θ lying between the upper endpoints of the 0.90 and 0.95 confidence intervals, etc. Of course this is logically incorrect, but it has powerful intuitive appeal. In the case where there are no nuisance parameters the confidence distribution is exactly Fisher's fiducial distribution for θ based on $\hat{\theta}$. Fisher proposed (1956) to use the fiducial distribution based on a primary sample as a prior distribution for the Bayesian analysis of a secondary sample. Lindley (1958) showed that this proposal leads to certain Bayesian incoherencies, setting off a vitriolic response from Fisher (1960).

We take notice of two things from these paragraphs. First, the definition of the confidence distribution itself (i.e., $\alpha_x(\theta)$) is purely frequentist and it does not involve any fiducial argument. In principle, piling up the boundaries of a set of confidence intervals of all levels for a parameter can typically give us a confidence distribution for the parameter. In particular, to rephrase the classical definition in the notation of this article and also in a slightly more general form, we have

(CL) *For every* α *in* $(0, 1)$, let $(-\infty, \tau_n(\alpha)]$ *be a* $100\alpha\%$ *lower-side confidence interval for a parameter* θ, *where* $\tau_n(\alpha) = \tau_n(\mathbf{x}, \alpha)$ *is continuous and increasing in* α *for each sample* $\mathbf{x}$. *Then,* $H_n(\cdot) = \tau_n^{-1}(\cdot)$ *is a confidence distribution for* θ.

Secondly, the interpretation of the confidence distribution is tangled with a fiducial reasoning, even though it is defined purely in the frequenstist domain. Historically, researchers have tried to interpret a confidence distribution as a sort of distribution of θ, which, typically, is the place where the fiducial argument is involved. Noting that the parameter θ is a fixed and non-random quantity in a frequentist setting, the interpretation (of treating a CD as an inherent distribution of a non-random quantity) is not possible, unless the fiducial reasoning (that θ is now also a random quantity) is invoked. In our view, this practice of adopting the fiducial argument has prevented the development of confidence distributions as a valuable statistical tool in frequentist inference. How to interpret a confidence distribution is one of the major departures of recent developments from the classical assertions. In the new interpretation, a confidence distribution is viewed as an estimator *for* the parameter of interest, instead of an inherent distribution *of* the parameter. This point will be further elaborated in this article.

2.2 A Modern Definition and Interpretation

The concept of confidence distribution has attracted a surge of renewed attention in recent years. We attribute the start of the renewed interest to Efron (1998), who "examined R.A. Fisher's influence on modern statistical thinking" and tried to "predict how Fisherian we can expect the 21st century to be". In the same article, Efron stated that a bootstrap distribution is "a distribution estimator" and "a confidence distribution". The idea of viewing a bootstrap distribution function as a (frequentist) distribution estimator of an unknown parameter has helped shape the viewpoint of treating a confidence distribution as a frquentist distribution estimator, instead as a fiducial distribution.

A confidence distribution is a "distribution estimator" and, conceptually, it is not different from a point estimator or a confidence interval. But it uses a sample-dependent distribution function on the parameter space to estimate a parameter of interest. As in the point estimation where any single point (a real value or a statistic) on the parameter space can in principle be used to estimate a parameter, any sample-dependent distribution function on the parameter space can in principle be used to estimate the parameter as well. But we impose some requirement on a confidence distribution to ensure that a statistical inference (e.g., point estimation, confidence interval, p-value, etc.) derived from it has desired frequentist properties. This practice is not different from the practice in the point estimation where we also set restrictions to ensure certain desired properties, such as unbiasedness, consistency, efficiency, etc.

The following definition is proposed and utilized in Schweder & Hjort (2002) and Singh *et al.* (2005, 2007). In the definition, Θ is the parameter space of the unknown parameter of interest θ, and $\mathcal{X}$ is the sample space corresponding to sample data $\mathbf{x} = \{x_1, \ldots, x_n\}$.

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

In non-technical terms, a confidence distribution is a function of both the parameter and the random sample, with two requirements. The first *R1)* is simply that, for each given sample, a confidence distribution should be a distribution function on the parameter space. The second *R2)* imposes a restriction to this sample-dependent distribution function so that inference based on it has desired frequentist properties. In essence, one requires that the distribution estimator is "coverage proper". Section 3 illustrates how to utilize the *R2)* requirement to extract information from a confidence distribution to make inference, including point estimation, confidence interval, p-value, etc.

Note that, when $\theta = \theta_0$ the true parameter value, *R2)* implies $H_n(\theta_0) \overset{sto}{=} 1 - H_n(\theta_0)$, but $H_n(\theta) \overset{sto}{\leq} 1 - H_n(\theta)$ for $\theta < \theta_0$, and $1 - H_n(\theta) \overset{sto}{\leq} H_n(\theta)$ for $\theta > \theta_0$ (see, Singh *et al.*, 2005). Here, "sto" means stochastic comparison between two random variables; i.e., $Y_1 \overset{sto}{\leq} Y_2$ means $P(Y_1 \leq t) \geq P(Y_2 \leq t)$ for all t. We may interpret this stochastic balancing equality at θ_0 and *R2)* to mean that the distribution estimator $H_n(\theta)$ contains "right" (or "balanced") information for making correct frequentist inference about the parameter.

Definition 1 is consistent with the classical definition of a confidence distribution. In particular, recall (CL) of Section 2.1 that a confidence distribution $H_n(\cdot)$ is defined as the inverse function of the upper limit of a lower-side confidence interval, i.e., $H_n(\cdot) = \tau_n^{-1}(\cdot)$. It follows that $\{\mathbf{x} : H_n(\theta) \leq \alpha\} = \{\mathbf{x} : \theta \leq \tau_n(\alpha)\}$, for any $\alpha \in (0, 1)$ and $\theta \in \Theta \subseteq \mathbb{R}$. Thus, at $\theta = \theta_0$, $\Pr\{H_n(\theta_0) \leq \alpha\} = \Pr\{\theta_0 \leq \tau_n(\alpha)\} = \alpha$ and $H_n(\theta_0) = \tau_n^{-1}(\theta_0)$ is $U[0, 1]$ distributed.

2.3 Illustrative Examples

The concept of a confidence distribution as defined in Section 2.2 subsumes and unifies a wide range of examples. We present next several illustrative examples, and more examples will be in Section 3.

2.3.1 Basic parametric examples

Example 1: *(Normal mean and variance)* Suppose that we have a sample

$$x_i \sim N(\mu, \sigma^2), \quad i = 1, \ldots, n. \tag{1}$$

When σ is known, it is clear that both

$$H_\Phi(\mu) = \Phi\left(\frac{\mu - \bar{x}}{\sigma/\sqrt{n}}\right) \quad \text{and} \quad H_t(\mu) = F_{t_{n-1}}\left(\frac{\mu - \bar{x}}{s/\sqrt{n}}\right) \tag{2}$$

satisfy the requirements in Definition 1; thus, both are confidence distributions for μ. Furthermore, $H_A(\mu) = \Phi(\sqrt{n}(\mu - \bar{x})/s)$ satisfies the definition of an asymptotic confidence distribution for μ, when $n \to \infty$; thus, it is an asymptotic confidence distribution for μ. Since $H_\Phi(\mu)$ and $H_A(\mu)$ are cumulative distribution functions of $N(\bar{x}, \sigma^2)$ and $N(\bar{x}, s^2)$, the uses of $H_\Phi(\mu)$ and $H_A(\mu)$ are equivalent to stating that we use $N(\bar{x}, \sigma^2)$ and $N(\bar{x}, s^2)$ *to estimate* μ, respectively.

When σ is not known, $H_\Phi(\mu)$ is no longer a confidence distribution for μ, since it involves the unknown σ and it violates the requirements in Definition 1. But we can still verify that $H_t(\mu)$ and $H_A(\mu)$ are a confidence distribution and an asymptotic confidence distribution for μ , respectively. Also, in this case, the sample-dependent distribution function on the parameter space of σ^2, $H_{\chi^2}(\theta) = 1 - F_{\chi^2_{n-1}}((n-1)s^2/\theta)$ is a confidence distribution for σ^2. Here, $F_{\chi^2_{n-1}}$ is the cumulative distribution function of the χ^2_{n-1}-distribution.

In Fisher's fiducial inference, $H_\Phi(\mu)$ is the fiducial distribution for the location parameter μ, when σ is known; and $H_t(\mu)$ is the fiducial distribution for μ, when σ is unknown. This is no coincidence and, in Section 3, we further demonstrate that many fiducial distributions satisfy the conditions required for being a confidence distribution. As in Fisher's fiducial development, a quantity known as a *pivotal function* plays an important role, especially in parametric examples like Example 1. In general, suppose $\psi(\mathbf{x}, \theta)$ is a pivot function and it is monotonic (without loss of generality, assume it is increasing) in θ. Then, the sample-dependent distribution function on the parameter space $H(\theta) = F\big(\psi(\mathbf{x}, \theta)\big)$ is a confidence distribution for θ, where F is the cumulative distribution function of the pivot quantity $\psi(\mathbf{x}, \theta)$. This device covers a large class of parameter examples.

2.3.2 Significant (p-value) functions

Consider a one-sided hypothesis test $K_0\colon \theta \leq b$ versus $K_1\colon \theta > b$, for a given b in the parameter space Θ. Let $p_n = p_n(b) = p_n(\mathbf{x}, b)$ be the p-value from a testing method. When b varies, this $p_n = p_n(b)$ forms a function on Θ, called a *significance function* or a *p-value function*; see, e.g., Fraser (1991). In most cases, this function $p_n(\cdot)$ is a cumulative distribution function, for every fixed sample $\mathbf{x}$. Also, at the true value $\theta = \theta_0$, $p_n(\theta_0)$, as a function of $\mathbf{x}$, is $U[0, 1]$-distributed or asymptotically $U[0, 1]$-distributed, provided that the tests achieve the nominal level of Type I error at least asymptotically. Thus, usually $p_n(\cdot)$ satisfies the requirements for a confidence distribution (or an asymptotic confidence distribution). The next example contains two such examples, one from a parametric test and the other from a non-parameteric test.

Example 2: (*p-value functions*) This first parametric test example is due to Fraser (2011). Consider the model $x = \mu + z$ where z has the standard extreme value (Gumbel) distribution with density $g(z) = e^{-z}e^{-e^{-z}}$ and cumulative distribution function $G(z) = e^{-e^{-z}}$. Suppose that we have a single sample x from this model. Based on a class of tests for one-sided hypothesis

K_0: $\mu \leq \theta$ versus K_1: $\mu > \theta$ for a given θ, we can get a p-value function

$$H_e(\theta) = \int_x^\infty g(z-\theta)dz = 1 - G(x-\theta) = 1 - e^{-e^{-(x-\theta)}}.$$

Clearly, $H_e(\theta)$ is a sample-dependent distribution function on the parameter space of the location parameter μ. Also, given $\theta = \mu$ the true parameter value, $H_e(\mu) = 1 - G(x-\mu) \sim U[0, 1]$, as a function of the random sample x. Thus, it is a confidence distribution for the location parameter μ.

Now let us turn to a different non-parametric testing problem. Denote by μ the center of symmetry of an unknown symmetric continuous distribution. Suppose there is an independently identically distributed sample $\mathbf{x}$ available from this population. A convenient confidence distribution in the context can be constructed through the p-value function

$$H_p(\theta) = p\text{-value of the one-sample signed rank test for hypothesis } K_0 : \mu \leq \theta \text{ vs. } K_1 : \mu > \theta. \quad (3)$$

It is not a hard exercise to show that such a function $H_p(\cdot)$ is typically non-decreasing, ranging from 0 to 1. But this $H_p(\cdot)$ is not an exact confidence distribution, since, although $H_p(\theta)$ is usually continuous in θ for the given sample $\mathbf{x}$, the distribution of $H_p(\theta)$ (as a function of $\mathbf{x}$) is discrete for a given θ. However, as $n \to \infty$, the discreteness vanishes and, when $\theta = \mu$ the true parameter value, $H_p(\mu)$ converges rapidly to $U[0, 1]$, under some mild conditions. Thus, $H_p(\cdot)$ is an asymptotic confidence distribution for the center parameter μ.

A p-value function $p_n(\cdot)$ is not a confidence distribution only in some unusual cases. For instance, a p-value function associated with a strictly conservative one-sided test does violate the requirements of being a confidence distribution. An example is the one-sided test for $\mu = \max\{\mu_1, \mu_2\}$ where μ_1 and μ_2 are population means of two separate studies. A conservative test based on Bonferroni adjustment would give rise to the p-value function $p(\theta) = 2\min\{p_1(\theta), p_2(\theta)\}$ for the hypotheses $K_0 : \mu \leq \theta$ vs. $K_1 : \mu > \theta$, where $p_1(\theta)$ and $p_2(\theta)$ are the corresponding p-value functions from individual studies. Clearly, $p(\cdot)$ violate the CD requirements, although the function can still be utilized to make inference (leading to conservative conclusions).

2.3.3 Bootstrap distributions

A bootstrap distribution is often an (asymptotic) confidence distribution. In any situations where one can construct a bootstrap distribution, one can construct a confidence distribution or an asymptotic confidence distribution. The following example is from Singh *et al.* (2001, 2007), who showed that a bootstrap distribution typically satisfies the definition of a confidence distribution.

Example 3: (*Nonparametric bootstrap*) Let $\hat\theta$ be an estimator of θ, and $\hat\theta^*$ be the estimator of $\hat\theta$ computed on a bootstrap sample. In the case when the limiting distribution of $\hat\theta$, properly normalized, is symmetric, the sampling distribution of $\hat\theta - \theta$ is estimated by the bootstrap distribution of $\hat\theta - \hat\theta^*$ (c.f., Efron & Tibshirani, 1994). Then, an asymptotic confidence distribution is given by

$$H_n(\theta) = 1 - P(\hat\theta - \hat\theta^* \leq \hat\theta - \theta|\mathbf{x}) = P(\hat\theta^* \leq \theta|\mathbf{x}),$$

which is also the raw bootstrap distribution of $\hat\theta$. In the case when the symmetry fails, the distribution of $\hat\theta - \theta$ can be estimated by the bootstrap distribution of $\hat\theta^* - \hat\theta$ (c.f., Efron &

Tibshirani, 1994). Then, the corresponding asymptotic confidence distribution is

$$H_n(\theta) = 1 - P(\hat{\theta}^* - \hat{\theta} \leq \hat{\theta} - \theta|\mathbf{x}) = P(\hat{\theta}^* \geq 2\hat{\theta} - \theta|\mathbf{x}).$$

We also consider another bootstrap-based confidence distribution by the bootstrap-t method, where the distribution of $(\hat{\theta} - \theta)/\widehat{SE}(\hat{\theta})$ is estimated by the bootstrap distribution of $(\hat{\theta}^* - \hat{\theta})/\widehat{SE}^*(\hat{\theta}^*)$. Here $\widehat{SE}^*(\hat{\theta}^*)$ is the estimated standard error of $\hat{\theta}^*$, based on the bootstrap sample. Such an approximation has so-called, second order accuracy (i.e., asymptotic error of order $O(n^{-1})$; see, e.g., Babu & Singh, 1983). The resulting asymptotic confidence distribution is

$$H_n(\theta) = 1 - P\left(\frac{\hat{\theta}^* - \hat{\theta}}{\widehat{SE}^*(\hat{\theta}^*)} \leq \frac{\hat{\theta} - \theta}{\widehat{SE}(\hat{\theta})}\middle|\mathbf{x}\right).$$

At $\theta = \theta_0$, this confidence distribution typically converges to $U[0, 1]$ at the rate of $O_p(n^{-1})$ and the confidence distribution is asymptotically second-order accurate.

2.3.4 Likelihood functions

Fiducial inference and likelihood functions are closely related. For instance, Fisher (1973) discussed extensively the connection between a likelihood function and a fiducial distribution, and Kendall & Stuart (1974) used a likelihood function to illustrate the concept of fiducial distribution. It is not surprising that a confidence distribution may also be closely related to a likelihood function. Welch & Peers (1963) and Fisher (1973) provided earlier accounts of likelihood-based confidence distributions in single parameter families. In particular, if we normalize a likelihood function curve with respect to its parameter(s) (provided that the normalizing is possible) so that the area underneath the curve is one, the normalized likelihood function curve is typically a density curve. Under some mild general conditions, Fraser & McDunnough (1984) showed that this normalized likelihood function is the density function of an asymptotic normal confidence distribution. Example 4 next is from Singh *et al.* (2007), who provided a formal proof that a profile likelihood function is proportional to an asymptotic normal confidence density for the parameter of interest.

Example 4: (*Profile likelihood function*) Suppose that there is an independently identically distributed sample of size n from a parametric distribution involving multiple parameters. Let $\ell_n(\theta)$ be the log profile likelihood function and $i_n^{-1} = -n^{-1}\ell_n''(\hat{\theta})$ with ni_n^{-1} being the observed Fisher information for a scalar parameter of interest θ. Denote by $\hat{\theta} = \text{argmax}_\theta \ell_n(\theta)$. Under the regularity conditions that ensure $\sqrt{n}(\hat{\theta} - \theta_0)/\sqrt{i_n} \to N(0, 1)$, plus some additional mild assumptions, Singh *et al.* (2007, theorem 4.1) proved that, for each given θ,

$$G_n(\theta) = H_n(\theta) + o_p(1), \quad \text{where } G_n(\theta) = \frac{\int_{(-\infty,\theta]\cap\Theta} e^{\ell_n(y)}dy}{\int_\Theta e^{\ell_n(y)}dy} \text{ and } H_n(\theta) = \Phi\left(\frac{\theta - \hat{\theta}}{\sqrt{i_n/n}}\right).$$

Since, at the true parameter value $\theta = \theta_0$, $H_n(\theta_0)$ converges to $U[0, 1]$, as $n \to \infty$. It follows that $G_n(\theta_0)$ converges to $U[0, 1]$, as $n \to \infty$. Thus, $G_n(\theta)$ is an asymptotic confidence distribution.

As functions of θ, the density function of $G_n(\theta)$ is proportional to $e^{\ell_n(\theta)}$. Thus, based on $G_n(\theta)$, we can provide an inference for θ that is asymptotically equivalent to that based on the profile likelihood function $e^{\ell_n(\theta)}$. From this observation and also Fraser & McDunnough (1994), we argue that a CD-based inference may subsume a likelihood inference in many occasions, especially when the standard asymptotic theory for the likelihood is involved. For practical purposes, normalizing a likelihood function with respect to the parameter(s) is a systematic scheme to obtain a confidence distribution, albeit it is usually an asymptotic normal confidence distribution under mild conditions.

In Section 7.1, we will further explore the connection between confidence distributions and likelihood functions. Specifically, we will discuss Efron's implied likelihood (Efron, 1993) and Schweder and Hjort's reduced likelihood (Schweder & Hjort, 2002), and illustrate how one can start with a confidence distribution to perform likelihood calculations.

2.3.5 Asymptotically third-order accurate confidence distributions

We end this section with two accounts of asymptotically third-order accurate (i.e., error of order $O(n^{-3/2})$) confidence distributions, one using a non-parametric construction and the other using a parametric construction. The techniques described here can be used to obtain versions of higher order improvement of Examples 3 and 4, and also to other settings.

Example 5: (*Third order accurate confidence distributions*) On the non-parametric side, there is a monotonic transformation, usually referred to as Hall's transformation (Hall, 1992), which combined with the bootstrap can produce a third order accurate confidence distribution. For a set of n sample $\mathbf{x}$ from an underlying distribution F, consider an asymptotic pivot of the form $T_n = (\hat{\theta} - \theta)/\widehat{SE}$, which admits Edgeworth expansion of the typical form

$$P(T_n \le y) = \Phi(y) + n^{-1/2}\big(a_F y^2 + b_F\big)\phi(y) + O(n^{-1}),$$

where a_F and b_F are smooth functions of the moments of F. Then, the monotonically transformed pivot $\psi(\mathbf{x}, \theta) = T_n + n^{-1/2}\big(\hat{a}_F T_n^2 + \hat{b}_F\big) + \frac{n^{-1}}{3}\hat{a}_F^2 T_n^3$ is free of the $n^{-1/2}$ term in its Edgeworth expansion (Hall, 1992). Here, $\hat{a}_F$ and $\hat{b}_F$ are the sample estimates of a_F and b_F. The bootstrap-based approximation of $P(\psi(\mathbf{x}, \theta) \le y)$ matches the $O(n^{-1})$ term, thus it is accurate up to an error of order $O(n^{-3/2})$. The third-order accurate confidence distribution would be given as

$$1 - \hat{G}_n(\psi(\mathbf{x}, \theta))$$

where $\hat{G}_n$ is bootstrap-based estimate of the cumulative distribution function of the pivot $\psi(\mathbf{x}, \theta)$. In the special case for the population mean μ, the monotonically transformed pivot is $\psi(\mathbf{x}, \mu) = t + \frac{\hat{\lambda}}{6\sqrt{n}}(2t^2 + 1) + \frac{1}{27n}\hat{\lambda}^2 t^3$, where $t = \sqrt{n}(\bar{x} - \mu)/s$, $\lambda = \mu_3/\sigma^3$, $\bar{x}$ is the sample mean, s^2 is sample variance and $\hat{\lambda}$ is a sample estimate of λ. It follows that a third-order correct confidence distribution for μ is given by $H_n(\mu) = 1 - \hat{G}_n\big(\psi(\mathbf{x}, \mu)\big)$.

On the parametric side, Reid & Fraser (2010) considered the standard exponential family with the log likelihood written in the form of

$$\ell(\gamma|\mathbf{x}) = \sum_{i=1}^{p} t_i(\mathbf{x}) s_i(\gamma) - C(\gamma) - h(\mathbf{x}).$$

Here, canonical variables $\mathbf{t}(\mathbf{x}) = (t_1(\mathbf{x}), \ldots, t_p(\mathbf{x}))^T$ are sufficient and $\mathbf{s}(\gamma) = (s_1(\gamma), \ldots, s_p(\gamma))^T$ are known functions of the vector of parameters $\gamma = (\theta, \Psi)$, where θ is a scaler parameter of interest and the nuisance parameter Ψ has $(p-1)$ components. Reid & Fraser (2010, p. 162) provided an explicit and general formula of a third-order accurate confidence distribution (or p -value function) for θ. This formula is accurate up to relative error of order $O(n^{-3/2})$ with the confidence distribution expressed as $H_{n(\theta)} = \Phi(r^*)$, where $r^* = r^*(\theta, \mathbf{x})$ is defined as

$$r^* = r + \frac{1}{r}\log\left(\frac{|q|}{|r|}\right), \quad \text{with } r = \text{sign}(\hat{\theta} - \theta)\big[2\ell(\hat{\gamma}_\theta)\big]^{\frac{1}{2}} \text{ and } q = \frac{|A|}{|S(\hat{\gamma})|}\frac{\{J(\hat{\gamma})\}^{1/2}}{|J_{\Psi\Psi}(\hat{\gamma}_\theta)|^{1/2}}.$$

In the formula, $\hat{\gamma}$ is the maximum likelihood estimator, $\hat{\gamma}_\theta = (\theta, \hat{\Psi}_\theta)$ is the constrained maximum likelihood estimator given a value of θ, $S(\gamma) = \{\partial/\partial\gamma\}\mathbf{s}(\gamma)$, $J_{\Psi\Psi}(\gamma)$ is the submatrix of $J(\gamma) = \{\partial^2/\partial\gamma\partial\gamma^T\}\ell(\gamma|\mathbf{x})$ corresponding to Ψ, and A is a $p \times p$ matrix whose first column is $\mathbf{s}(\hat{\gamma}) - \mathbf{s}(\hat{\gamma}_\theta)$ and the remaining $(p-1)$ columns are the last $(p-1)$ columns of $S(\hat{\gamma})$.

3 Confidence Distribution, Bootstrap, Fiducial, and Bayesian Approaches

3.1 CD-Random Variable, Bootstrap Estimator, and Fiducial-less Interpretation

For each given sample $\mathbf{x}$, $H_n(\cdot)$ is a cumulative distribution function on the parameter space. We can construct a random variable ξ defined on $\mathcal{X} \times \Theta$ (with a suitable probability measure) such that, conditional on the sample data, ξ has the distribution $H_n(\cdot)$. For example, let U be a $U[0, 1]$ random variable that is independent of sample data $\mathbf{x}$. Then, given $\mathbf{x}$, $\xi = H_n^{-1}(U)|\mathbf{x} \sim H_n(\cdot)$. We call this random variable ξ a *CD-random variable* (see, e.g., Singh *et al.,* 2007).

Definition 2: We call $\xi = \xi_{H_n}$ a CD-random variable associated with a confidence distribution $H_n(\cdot)$, if the conditional distribution of ξ given the data $\mathbf{x}$ is $H_n(\cdot)$.

Unlike the case of fiducial inference, the CD-random variable ξ is not a “random parameter” (a random version of θ). Rather it may be viewed as a CD-randomized estimator of θ_0. As an estimator, ξ is median unbiased, i.e., $P_{\theta_0}(\xi \le \theta_0) = E_{\theta_0}\{H_n(\theta_0)\} = \frac{1}{2}$.

To better understand the concept of the CD-random variable, we explore its close association with a bootstrap estimator. Consider first the simple normal case of Example 1 with a known variance σ^2. The mean parameter μ is estimated by the distribution $N(\bar{x}, \sigma^2)$. A CD-random variable ξ follows $\xi|\bar{x} \sim N(\bar{x}, \sigma^2)$ and we have

$$\left.\frac{\xi - \bar{x}}{\sigma/\sqrt{n}}\right|\bar{x} \sim \left.\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right|\mu \qquad (\text{both} \sim N(0, 1)).$$

This statement is exactly the same as the key justification for bootstrap, replacing ξ by a bootstrap sample mean $\bar{x}^*$. So in essence ξ is the same as the bootstrap estimator $\bar{x}^*$. Now, consider more generally a typical setup for a bootstrap procedure, as in Example 3. Let $\mathbf{x}$ be the data, $\hat{\theta}$ be an estimator of θ and $\hat{\theta}^*$ be the estimator of $\hat{\theta}$ computed on a bootstrap sample. As illustrated in Example 3, in the case when the sampling distribution of $\hat{\theta} - \theta$ is estimated by the bootstrap distribution of $\hat{\theta} - \hat{\theta}^*$, $H_n(\theta) = 1 - P(\hat{\theta} - \hat{\theta}^* \le \hat{\theta} - \theta|\mathbf{x}) = P(\hat{\theta}^* \le \theta|\mathbf{x})$ turns out to be an asymptotic confidence distribution for θ. In this case, given $\mathbf{x}$,

$$\xi = \hat{\theta}^* \,\big|\, \mathbf{x} \sim H_n(\cdot).$$

Thus, $\xi = \hat{\theta}^*$ is a CD-random variable associated with $H_n(\theta)$. In the case when the sampling distribution of $\hat{\theta} - \theta$ is estimated by the bootstrap distribution of $\hat{\theta}^* - \hat{\theta}$, we have that $H_n(\theta) = 1 - P(\hat{\theta}^* - \hat{\theta} \le \hat{\theta} - \theta|\mathbf{x}) = P(\hat{\theta}^* \ge 2\hat{\theta} - \theta|\mathbf{x})$ is an asymptotic confidence distribution and, given $\mathbf{x}$,

$$\xi = 2\hat{\theta} - \hat{\theta}^* \,\big|\, \mathbf{x} \sim H_n(\cdot).$$

Thus, $\xi = 2\hat{\theta} - \hat{\theta}^*$ is a CD-random variable. Clearly, when a bootstrap procedure applies, the bootstrap estimator $\hat{\theta}^*$ is closely related to a CD-random variable. Loosely speaking, the

CD-random variable ξ is in essence the same as a bootstrap estimator. This close connection between the CD-random variable and a bootstrap estimator may inspire a possible view of treating the concept of confidence distribution as an extension of a bootstrap distribution, albeit the concept of confidence distribution is much broader.

The connection to bootstrap mentioned above and the well-developed theory of bootstrap distributions can help us understand inference procedures involving confidence distributions and develop new methodologies. For instance, the CD-random variable may also be utilized to offer a new simulation mechanism, which can broaden applications of the standard bootstrap procedures, especially when only a data summary is available. The CD-random variable ξ is in essence a bootstrap estimator, but its interpretation is not limited to being just a bootstrap estimator. This freedom allows us to use ξ more liberally, which in turn allows us to develop more flexible approaches and simulations. See Section 7.3 for such examples. Also, just like a bootstrap estimator, the CD-random variable can be extended to the case of parameter vectors. This clue may help us develop the concept of a multi-dimensional confidence distribution. See Section 8 for further discussions on the topic of multi-dimensional confidence distributions.

Viewing a CD-random variable as a quantity similar to a bootstrap estimator can also help us further clarify the concept of a confidence distribution by avoiding the confusing fiducial interpretation that the parameter is both a fixed and a random quantity. It is well known that a fiducial distribution is not an ordinary probability distribution function in the frequentist sense (e.g., Kendall & Stuart, 1974, chapter 21). Since a confidence distribution has been historically interpreted as a fiducial distribution, it is also a common assertion that a confidence distribution is not a proper distribution in the frequentist sense (see, e.g., Cox, 2006, p. 66; Schweder & Hjort, 2002, p. 310, 328). We think this assertion about confidence distributions could perhaps be modified, if we treat and interpret a confidence distribution as a purely frequenist concept without any fiducial reasoning (as we do in the current context). Our view is that, like a bootstrap distribution, a confidence distribution is an ordinary probability distribution function for each given sample. However, since we no longer view a confidence distribution as an inherent distribution of θ, we can *not* manipulate it as if it is a distribution of θ to automatically reach a conclusion. For instance, we may manipulate a confidence distribution of θ, for example using a variable transformation $g(\theta)$ of θ, to get another sample-dependent distribution function on the corresponding parameter space. In general, this new sample-dependent distribution function may *not* be a confidence distribution for $g(\theta)$, unless $g(\cdot)$ is monotonic. For an asymptotic confidence distribution, the monotonicity is only needed locally in a neighborhood of the true θ value thus it covers a large class of smooth function $g(\cdot)$. This equivariant property with respect to a monotone transformation of θ is the same as that of a confidence interval.

The following example was brought to our attention by Professor David R. Cox in our email communications. We think it provides a good example to illustrate the above viewpoint.

Example 6: (Ratio of two normal means) Suppose we have two sets of independent normal samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. For simplicity, let us assume that both σ_1^2 and σ_2^2 are known. From Example 1, $N(\bar{x}_1, \sigma_1^2/n_1)$ and $N(\bar{x}_2, \sigma_2^2/n_2)$ are confidence distributions for μ_1 and μ_2, respectively. In fact, they are also the fiducial distributions of μ_1 and μ_2. Here, $\bar{x}_i$ and n_i, $i = 1, 2$, are the sample means and sample sizes of the two samples, respectively. In the fiducial inference, $N(\bar{x}_1, \sigma_1^2/n_1)$ and $N(\bar{x}_2, \sigma_2^2/n_2)$ are treated as inherent distributions of μ_1 and μ_2. Thus, if we treat them as the usual distributions of two independent random variables, a distribution of $\delta = \mu_1/\mu_2$ can be easily obtained by the standard manipulation, assuming the true $\mu_2 \neq 0$. However, this distribution of δ is not good for exact fiducial inference, and this discrepancy is typically explained in the classical literature by the assertion that fiducial distributions are not and can not be manipulated as proper distribution functions. It remains unclear how to manipulate the two normal distributions to obtain the (unique) fiducial distribution of δ, in the classical sense.

In the CD-inference, we do not encounter the same interpretation problem, as we treat $N(\bar{x}_1, \sigma_1^2/n_1)$ and $N(\bar{x}_2, \sigma_2^2/n_2)$ as distribution estimators for μ_1 and μ_2 which happen to have the property of "proper coverage probability". We can manipulate $N(\bar{x}_1, \sigma_1^2/n_1)$ and $N(\bar{x}_2, \sigma_2^2/n_2)$ to get one or more a sample-dependent distribution functions on the parameter space of $\delta = \mu_1/\mu_2$. For example, among other possibilities, we may use a direct result from the standard manipulation of the two independent normal confidence distributions to estimate δ, $H_r(\delta) = \Phi\big\{(\delta - \bar{x}_1/\bar{x}_2)/s_r\big\}$, where $s_r^2 = (1/\bar{x}_2^2)\sigma_1^2/n_1 + (\bar{x}_1^2/\bar{x}_2^4)\sigma_2^2/n_2$. It is true that $H_r(\delta)$, a sample-dependent distribution function on $(-\infty, \infty)$, has lost the exact coverage property. But this phenomenon is not different from what we see in some point estimation practices. For instance, the sample means $\bar{x}_1$ and $\bar{x}_2$ are unbiased estimators of μ_1 and μ_2, and their ratio $\bar{x}_1/\bar{x}_2$ is still a decent point estimator of $\delta = \mu_1/\mu_2$ but no longer unbiased. When both $n_i \to \infty$, $\bar{x}_1/\bar{x}_2$ is a consistent (also "asymptotically unbiased") estimator of $\delta = \mu_1/\mu_2$. Similarly, $H_r(\delta)$ still offers proper coverage asymptotically when both $n_i \to \infty$ and thereby is still an asymptotic confidence distribution for $\delta = \mu_1/\mu_2$.

3.2 Confidence Distribution, Fiducial Distribution, and Belief Function

Both the approach of confidence distribution and Fisher's fiducial approach share a common goal to provide a "distribution estimator" for a parameter of interest. As stated in a review article by Hannig (2009), a fiducial inference can be "viewed as a procedure that obtains a measure on a parameter space"; also "statistical methods designed using the fiducial reasoning have typically very good statistical properties as measured by their repeated sampling (frequentist) performance". The same two considerations are those behind the modern definition of confidence distribution, expressed as the two requirements in Definition 1. It is not surprising that many fiducial distributions meet the conditions required for being a confidence distribution.

The next example is a fiducial distribution used by Kendall & Stuart (1974, example 21.2, p. 144) to illustrate the fiducial method for "a non-symmetrical sampling distribution". We use it to demonstrate that the fiducial distribution satisfies the requirements as a confidence distribution.

Example 7: (Scale parameter of gamma distribution) Consider a Gamma(p_0, θ) distribution, $f(x|\theta, p_0) = \{x^{p_0-1}e^{-x/\theta}\}/\{\theta^{p_0}\Gamma(p_0)\}$, where the scale parameter θ is the unknown parameter of interest and the shape parameter p_0 is known. Here, $\Gamma(\cdot)$ is the gamma function. Kendall & Stuart (1974, example 21.2, p. 144) provided the fiducial distribution for θ with a density $\big(n\bar{x}/\theta\big)^{np_0}\big\{e^{-n\bar{x}/\theta}/\,\Gamma(np_0)\big\}(1/\theta)d\theta$, or in a cumulative distribution function form,

$$H_n(\theta) = \int_0^\theta \left(\frac{n\bar{x}}{t}\right)^{np_0} \frac{e^{-n\bar{x}/t}}{\Gamma(np_0)t}dt,$$

where $\bar{x}$ is the sample mean. Clearly, this $H_n(\theta)$ is a cumulative distribution function on the parameter space $\Theta = (0, \infty)$. Also, by the variable transformation $\tilde{t} = n\bar{x}\theta/t$ in the integration, the function $H_n(\theta)$ can be re-expressed as

$$H_n(\theta) = \int_{n\bar{x}}^\infty \left(\frac{\tilde{t}}{\theta}\right)^{np_0} \frac{e^{-\tilde{t}/\theta}}{\Gamma(np_0)\tilde{t}}d\tilde{t} = 1 - F_{\Gamma(np_0,\theta)}(n\bar{x}),$$

where $F_{\Gamma(np_0,\theta)}(\cdot)$ is the cumulative distribution function of a Gamma(np_0, θ) distribution. At the true parameter value $\theta = \theta_0$, $n\bar{x}$ follows Gamma(np_0, θ_0) distribution. It follows immediately that

$H_n(\theta_0) = 1 - F_{\Gamma(np_0,\theta_0)}(n\bar{x}) \sim U[0, 1]$. Thus, by Definition 1, the fiducial distribution $H_n(\theta)$ is a confidence distribution for the scale parameter θ.

See, also, Wong (1993, 1995) for formulas of confidence distributions for unknown parameters of a Gamma distribution, when both the scale and shape parameters are unknown.

Although fiducial distributions are confidence distributions in many examples, the developments of confidence distributions, especially the recent efforts, are not a part of any fiducial developments. The concept of a confidence distribution by its definition is very broad, and the definition itself does not provide a standard procedure how it can be constructed (e.g., Definition 1, or in other forms). Singh & Xie (2011) described the relation between the concepts of confidence and fiducial distributions using an analogy in point estimation: a consistent estimator versus a maximum likelihood estimator (MLE). A confidence distribution is analogous to a consistent estimator which is defined to ensure a certain desired property for inference; a fiducial concept is analogous to an MLE which provides a standard procedure to find an estimator which often happens to possess desirable frequentist properties. A consistent estimator does not have to be an MLE; but, under some regularity conditions, the MLE typically has consistency and the MLE method thus provides a standard procedure to obtain a consistent estimator. In the context of distribution estimation, a confidence distribution does not have to be a fiducial distribution or involve any fiducial reasoning. But, under suitable conditions, a fiducial distribution may satisfy the required "good properties (as measured by coverage and length of confidence intervals based on the fiducial distribution)" (c.f., Hannig, 2009), which thereby establishes it as a confidence distribution.

There has been also renewed interest in fiducial inference and its extensions, notably the recent developments on "generalized fiducial inference" by S. Weerahandi, J. Hannig, H. Iyer, and their colleagues and the developments of belief functions under Dempster-Shafer theory by Arthur Dempster, Glenn Shafer, Chuanhai Liu, and others. See, e.g., review articles by Hannig (2009), Dempster (2008), and Martin *et al.* (2010). These developments, together with the new developments of confidence distributions, represent an emerging new field of research on distributional inferences. A confidence distribution as defined, interestingly, can fit into the framework of generalized fiducial inferences (c.f., Hannig, 2009), since a fiducial distribution $Q(U, \mathbf{x})$ can be obtained by solving a generalized fiducial equation $H(\cdot) = U$ with $H(\cdot) = H(\mathbf{x}, \cdot)$ being a confidence distribution and U being a $U[0, 1]$-distributed random variable, provided it exits. However, the development of confidence distributions are distinct from the developments of fiducial and belief functions in that it is developed strictly within the frequentist domain and resides entirely within the frequentist logic, without involving any new theoretical framework such as fiducial reasoning or Dempster-Shafer theory. Contrary to the belief function, the CD-based approaches are easy to implement and can be directly related to a vast collection of examples used in classical and current statistical practices. Nevertheless, fiducial inference provides a systematic way to obtain a confidence distribution. Additionally, its development provides a rich class of literature for CD inference, and can generate new insight and development beyond frequentist inference.

3.3 Confidence Distribution and Bayesian Inference

A Bayes posterior can provide credible intervals that often have no assurance of (exact) frequentist coverage. Fraser (2011) expounded the basic fact that, in the presence of linearity in a model (e.g., a location model), a Bayes posterior with respect to the flat prior can be exactly a confidence distribution, which appears to be a mathematical coincidence. Fraser (2011) also demonstrated how the departure from the linearity induces the departure of a posterior from being a confidence distribution, in a proportionate way. But on asymptotic grounds, researchers have treated Bayes credible intervals as confidence intervals in many applications. As is well known, when more and more data accumulates, the prior effect fades away and a posterior density typically converges to a normalized likelihood function; see, e.g., LeCam (1953, 1958), Johnson (1967) among others. Since a normalized likelihood function is typically a confidence density in the limit, we argue that a posterior distribution can be treated as a (first-order) asymptotic confidence distribution as well. Because in many situations a Bayesian approach, especially an objective Bayesian approach, can often obtain a posterior distribution whose credible intervals have exact or asymptotic frequentist coverages, we also view the Bayes method as one of the machineries that can potentially produce confidence distributions.

Next Example 8 is from Xie *et al.* (2011), who argued from the angle of a confidence distribution that an informative normal prior and its corresponding posterior after incorporating a normal sample can sometimes be viewed as frequentist confidence distributions. The argument helped bring the conventional normal-model-based Bayesian meta-analysis under a unifying meta-analysis framework in Xie *et al.* (2011). The example is a mathematical coincidence, hinged on the normal assumption.

Example 8: (Informative prior and posterior distribution) Suppose $\pi(\theta) \sim N(\mu_0, \sigma_0^2)$ is an informative prior of a parameter θ of interest. Assume this informative prior is formed on the basis of extensive prior information of θ from past results of the same or similar experiments. Suppose Y_0 is a normally distributed summary statistic from these past experiments, with a realization $Y_0 = \mu_0$ and an observed variance var$(Y_0) = \sigma_0^2$, respectively. If we denote by $\mathcal{X}_0$ the sample space of the past experiments and by Θ the parameter space of θ, we can show by the definition that $H_0(\theta) = \Phi((\theta - Y_0)/\sigma_0)$ is a confidence distribution on $\mathcal{X}_0 \times \Theta$. Thus, we consider $H_0(\theta) = \Phi((\theta - \mu_0)/\sigma_0)$ as a distribution estimate from the past experiments. That is, the past experiments produced $N(\mu_0, \sigma_0^2)$ as a distribution estimate for θ.

Now, suppose that we have a set of observed sample from $N(\theta, \sigma^2)$. Based on the sample data and as in Example 1, we can obtain a confidence distribution for θ, $H_\Phi(\theta) = \Phi\big(\sqrt{n}(\theta - \bar{x})/\sigma\big)$, when σ^2 is given, or an asymptotic confidence distribution $H_A(\theta) = \Phi\big(\sqrt{n}(\theta - \bar{x})/s\big)$, regardless of whether σ^2 is known or not. Here, $\bar{x}$ is the sample mean and s^2 is the sample variance. Using a recipe for combining confidence distributions (see, e.g., Singh *et al.*, 2005 and Xie *et al.*, 2011, also Section 6 of this article), we can combine the prior confidence distribution $H_0(\theta)$ with the confidence distribution from the data, either $H_\Phi(\theta)$ or $H_A(\theta)$, to obtain a combined function (defined on $\{\mathcal{X}_0 \times \mathcal{X}\} \times \Theta$),

$$H_\Phi^{(c)} = \Phi\left(\frac{\theta - \hat{\mu}_\Phi}{\hat{\sigma}_\Phi}\right) \quad \text{or} \quad H_A^{(c)} = \Phi\left(\frac{\theta - \hat{\mu}_A}{\hat{\sigma}_A}\right),$$

where $\hat{\mu}_\Phi = (\mu_0/\sigma_0^2 + \bar{x}/\sigma^2)/(1/\sigma_0^2 + 1/\sigma^2)$, $\hat{\sigma}_\Phi^2 = (1/\sigma_0^2 + 1/\sigma^2)^{-1}$ and $\hat{\mu}_A = (\mu_0/\sigma_0^2 + \bar{x}/s^2)/(1/\sigma_0^2 + 1/s^2)$, $\hat{\sigma}_A^2 = (1/\sigma_0^2 + 1/s^2)^{-1}$. By the definition, we can show that $H_\Phi^{(c)}(\theta)$ is a confidence distribution for θ when σ^2 is known and $H_A^{(c)}(\theta)$ is an asymptotic confidence distribution for θ regardless of whether σ^2 is known or not. In other words, the confidence distributions $N(\hat{\mu}_\Phi, \hat{\sigma}_\Phi^2)$ or $N(\hat{\mu}_A, \hat{\sigma}_A^2)$ can be used to estimate the mean θ. The confidence distribution $N(\hat{\mu}_\Phi, \hat{\sigma}_\Phi^2)$ is identical in form to the posterior distribution when σ^2 is known, and the asymptotic confidence distribution $N(\hat{\mu}_A, \hat{\sigma}_A^2)$ is asymptotically equivalent to the posterior distribution obtained using the Bayes formula, regardless of whether σ^2 is known or not.

In general, an approach based on a confidence distribution is a frequentist procedure, and it inherits properties of frequentist procedures. Wasserman (2007) re-examined Efron's 1986 paper titled “Why not everyone is a Bayesian?”. He summarized from Efron (1986) four aspects of difference between the frequentist and Bayesian approaches: “easy of use”, “model building”, “division of labor”, and “objectivity”. He also added the extra item of “randomization”, and stated that “things haven’t changed much” since Efron (1986), “with the possible exception of (2)” on model building. Generally speaking, modelling and making inference based on confidence distributions are different from a Bayesian procedure in these four or five aspects. Here, we examine only the aspect “division of labor”, which is highlighted in several developments on confidence distributions.

Under the item “division of labor”, Wasserman (2007) stated that “The idea that statistical problems do not have to be solved as one coherent whole is anathema to Bayesian but is liberating for frequentists. To estimate a quantile, an honest Bayesian needs to put a prior on the space of all distributions and then find the marginal posterior. The freuqentist need not care about the rest of the distribution and can focus on much simpler tasks”. This aspect underscores a difference on how to handle nuisance parameters between a Bayesian and a frequentist approach, and some of the CD-based developments have indeed keyed on the feature of not involving nuisance parameters. For instance, the main selling point of the CD-based likelihood calculations in Efron (1993) and Schweder & Hjort (2002), as well as the development of Singh & Xie (2012), is the avoidance of nuisance parameters. The “discrepant posterior phenomenon” reported in Xie *et al.* (2013) in a binomial clinical trial from a real pharmaceutical application can also shed further insight and generate discussion on this aspect of division of labor. A concrete example of this discrepant posterior phenomenon and Xie *et al.* (2013) will be further discussed in details later in Section 6.2.

Finally, it is beneficial for the field of statistics to have both Bayesian and frequentist methods. There are bridges between these different schools of inference, and the Bayesian and frequentist methods can enrich themselves from taking inspiration from each other’s strong suit. For instance, combining evidence from disparate sources of information has long been considered a weak point of frequentist theory, while being a strong suit of Bayesian analysis. The flexibility of a Bayesian approach in this and other areas have helped stimulate the development of CD-based methods in related areas. Nevertheless, the CD development may also assist the development of objective Bayesian approaches. For instance, a CD-based approach could potentially be a good way to obtain an objective prior where we can use a variety of frequentist tool kits, including robust inference. As Kass (2011) stated, statistics has moved beyond the frequentist-Bayesian controversies of the past. It is our hope that evolving understanding and continuing developments of confidence and fiducial distributions will stimulate further developments that can enrich statistical theory and applications, and bridge the gaps among the Bayesian, fiducial, and frequentist philosophies.

4 Inferences Using a Confidence Distribution

Just like a Bayesian posterior that contains a wealth of information for any type of Bayesian inference, a confidence distribution contains a wealth of information for constructing any type of frequentist inference. Discussions on how to make inference using a confidence distribution have been scattered around in many publications, for examples, in Fisher (1973), Cox & Hinkley (1974), Mau (1988), Fraser (1991), Barndorff-Nielsen & Cox (1994), Schweder & Hjort (2002), Singh *et al.* (2001, 2005, 2007), Coudin & Dufour (2004), among others. Here, in a more systematic fashion and under a general setting, we illustrate several aspects of making inference using a confidence distribution. A simple graphical illustration of the main message is provided in Figure 1 with mathematical details presented in the remaining of this section.

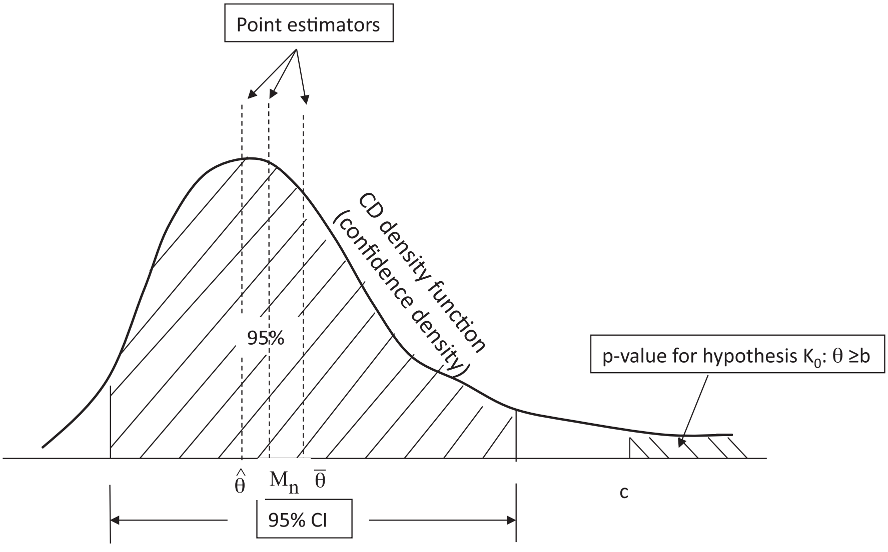

Figure 1. *The plot is a graphical illustration on making inference using a confidence distribution, including examples of point estimators (mode $\hat{\theta}$, median M_n and mean $\bar{\theta}$), a level 95% confidence interval and a one-sided p-value.*

4.1 Confidence Interval

As mentioned before, a confidence distribution is a sample-dependent distribution function that can represent confidence intervals of all levels for a parameter of interest. It is evident from requirement R2) in Definition 1 that the intervals $(-\infty, H_n^{-1}(1-\alpha)]$ and $[H_n^{-1}(\alpha), +\infty)$ provide level $100(1-\alpha)\%$ one-sided confidence intervals for the parameter of interest θ, for any $\alpha \in (0, 1)$. Also, $(H_n^{-1}(\alpha_1), H_n^{-1}(1-\alpha_2))$ is a level $100(1-\alpha_1-\alpha_2)\%$ confidence interval for the parameter θ, for any $\alpha_1 > 0, \alpha_2 > 0$, and $\alpha_1 + \alpha_2 < 1$. The same is true for an asymptotic confidence distribution, where the confidence level is achieved in limit.

4.2 Point Estimation

A confidence distribution, say $H_n(\cdot)$, can also provide us a point estimator for the parameter of interest. Natural choices of point estimators of the parameter θ given $H_n(\cdot)$, include the median $M_n = H_n^{-1}(1/2)$, the mean $\bar{\theta}_n = \int_{t\in\Theta} t\, dH_n(t)$, and the mode $\hat{\theta}_n = \arg\max_\theta h_n(\theta)$ of the confidence density function $h_n(\theta) = \frac{d}{d\theta}H_n(\theta)$. Under some modest conditions, we can prove that these three point estimators are consistent estimators. In particular, we assume that

(A) *For any* ϵ, $0 < \epsilon < \frac{1}{2}$, $L_n(\epsilon) = H_n^{-1}(1-\epsilon) - H_n^{-1}(\epsilon) \to 0$, in probability, as the sample size $n \to \infty$.

This condition is equivalent to (see a proof in Xie *et al.*, 2011)

(A′) *For any fixed* $\delta > 0$, $H_n(\theta_0 - \delta) \to 0$ and $H_n(\theta_0 + \delta) \to 1$, in probability, as $n \to \infty$.

We interpret these conditions as: as the sample size n increases, the (density) mass of the confidence distribution $H_n(\theta)$ becomes more and more concentrated around the true value θ_0. We have the following theorem regarding the three point estimators. A proof of the theorem can be found in Singh *et al.* (2007).

THEOREM 1. *Under Condition (A), as $n \to \infty$,* (i) *M_n is consistent estimator and, moreover, if* $L_n(\epsilon) = O_p(a_n)$, for a non-negative $a_n \to 0$, then $M_n - \theta_0 = O_p(a_n)$; *(ii) if $r_n = \int_{t\in\Theta} t^2 dH_n(t)$ is bounded in probability, then $\bar{\theta}_n$ is a consistent estimator; and (iii) if there exists a fixed $\epsilon > 0$* such that $P\big(\hat{\theta}_n \in \big[H_n^{-1}(\epsilon), H_n^{-1}(1-\epsilon)\big]\big) \to 1$, then $\hat{\theta}_n$ *is a consistent estimator.*

Furthermore, the median M_n is a median unbiased estimator with $P_{\theta_0}(M_n \le \theta_0) = P_{\theta_0}(1/2 \le H_n(\theta_0)) = 1/2$. The mean $\bar{\theta}_n$ is the frequentist analog of Bayesian estimator of θ under the usual squared error loss. The mode $\hat{\theta}_n$ matches with the maximum likelihood estimator if $h_n(\theta)$ is from a normalized likelihood function. If $H_n(\cdot)$ is constructed from a hypothesis test, the point estimators may also be related to a Hodges-Lehmann-type estimator (see, e.g., Coudin & Dufour, 2004).

4.3 *Hypothesis Testing*

From a confidence distribution, we can make inference for various hypothesis testing problems. Fraser (1991) developed several results on this topic through significance (p-value) functions. Singh *et al.* (2001, 2007) also provided several results under the framework of confidence distributions. See, also, e.g., Birnbaun (1961), Cox & Hinkley (1974), Mau (1988), Barndorff-Nielsen & Cox (1994), Schweder & Hjort (2002) for discussions on related topics.

Let us start with a one-sided test $K_0 : \theta \in C$ versus $K_1 : \theta \in C^c$, where C is an interval of the type $(-\infty, b]$ or $[b, \infty)$, for a given b, and C^c is the complementary set of C. We like to measure the "support" that $H_n(\cdot)$ lends to C, i.e., the probability content of C under $H_n(\cdot)$,

$$p_s(C) = \int_{\theta\in C} dH_n(\theta) = H_n(C), \tag{4}$$

which is also interpreted as a probability of belief under the framework of fiducial inference (e.g., Kendall & Stuart (1974), chapter 21). We use the term "support", instead of"belief", to highlight that the use of the confidence distribution does not involving any fiducial reasoning. If the support on C is high, we choose C and, if it is low, we choose C^c. Specifically, we reject the null hypothesis K_0 if the support on C is less than α (or the support on C^c is greater than $1-\alpha$), i.e., the rejection region is:

$$\{\mathbf{x} : p_s(C) \le \alpha\} \quad \text{or, equivalently,} \quad \{\mathbf{x} : p_s(C^c) \ge 1-\alpha\}. \tag{5}$$

We can prove that $\sup_{\theta\in C} P_\theta\big(p_s(C) \le \alpha\big) = \alpha$. This is because, for any $\theta \in C = (-\infty, b]$, $P_\theta\big(p_s(C) \le \alpha\big) \le P_\theta\big(p_s((-\infty, \theta]) \le \alpha\big) = \alpha$, where the last equation holds because $p_s((-\infty, \theta]) = H_n(\theta) \sim U[0, 1]$ under the assumption that θ is the true parameter, and also an equality holds throughout when the true $\theta = b$. The same argument also applies to the right-sided test case with $C = [b, \infty)$. Thus, the rejection region (5) corresponds to a level α test.

In the case of the one-sided tests with K_0 being of the type $(-\infty, b]$ or $[b, \infty)$, the value of $p_s(C) = H_n(C)$ is usually the conventional p-value. This is illustrated in the following example.

Example 9. Consider a set of sample from $N(\mu, \sigma^2)$, where both μ and σ^2 are unknown. As stated in Examples 1, a confidence distribution for μ is $H_t(\mu) = F_{t_{n-1}}\big(\sqrt{n}(\mu - \bar{x})/s\big)$. For a one-sided test $K_0 : \mu \le b$ versus $K_1 : \mu > b$, its support on the null set $C = (-\infty, b]$ is

$$p_s(C) = p_s((-\infty, b]) = H_t(b) = F_{t_{n-1}}\big(\sqrt{n}(b - \bar{x})/s\big).$$

This is exactly the p-value using the t-test for the one-sided test $K_0 : \mu \le b$ versus $K_1 : \mu > b$.

Now, let us move to a two-sided point hypothesis $K_0 : \theta = b$ versus $K_1 : \theta \neq b$, for a fixed b. In this case with the set $C = \{b\}$ being a singleton, we like to measure the supports of its two alternative sets $p_s(C_{lo}^c)$ and $p_s(C_{up}^c)$, where C_{lo}^c is the complementary set of $C_{lo} = (-\infty, b]$ and C_{up}^c is the complementary set of $C_{up} = [b, \infty)$. We define the rejection region as $\left\{\mathbf{x} : 2\max\{p_s(C_{up}^c), p_s(C_{lo}^c)\} \geq 1-\alpha\right\}$ or, equivalently,

$$\{\mathbf{x} : 2\min\{p_s(C_{lo}), p_s(C_{up})\} \leq \alpha\} = \{\mathbf{x} : 2\min\{H_n(b), 1 - H_n(b)\} \leq \alpha\}. \tag{6}$$

It is immediate that, under K_0 with the true $\theta = b$, $2\min\{p_s(C_{lo}), p_s(C_{up})\} = 2\min\{H_n(b), 1 - H_n(b)\} \sim U[0, 1]$, since $H_n(b) \sim U[0, 1]$ by the definition of a confidence distribution. Thus, $P_{K_0:\theta=b}\big(2\min\{p_s(C_{lo}), p_s(C_{up})\} \leq \alpha\big) = P_{K_0:\theta=b}\big(2\min\{H_n(b), 1 - H_n(b)\} \leq \alpha\big) = \alpha$ and the rejection region (6) corresponds to a level α test. Again, the value of $2\min\{p_s(C_{lo}), p_s(C_{up}) = 2\min\{H_n(b), 1 - H_n(b)\}$ is in fact the conventional p-value, as illustrated in the next example.

Example 10. (Continues from Example 9). If the null hypothesis is $K_0 : \mu = \mu_0$ versus $K_1 : \mu \neq \mu_0$, we have

$$\begin{aligned} 2\min\{p_s(C_{lo}), p_s(C_{up})\} &= 2\min\{H_t(\mu_0), 1 - H_t(\mu_0)\} \\ &= 2\min\left\{F_{t_{n-1}}\left(\frac{\mu_0-\bar{x}}{s/\sqrt{n}}\right), 1 - F_{t_{n-1}}\left(\frac{\mu_0-\bar{x}}{s/\sqrt{n}}\right)\right\}, \end{aligned}$$

This expression agrees with the standard p-values based on the two-sided t test.

For the purpose of presentation, we summarize the above results in a formal theorem.

THEOREM 2. *(i) For the one-sided test $K_0 : \theta \in C$ versus $K_1 : \theta \in C^c$, where C is an interval of the type of $(-\infty, b]$ or $[b, \infty)$, we have* $\sup_{\theta\in C} P_\theta\big(p_s(C) \leq \alpha\big) = \alpha$ and $p_s(C) = H_n(C)$ is the corresponding *p-value of the test. (ii) For the singleton test $K_0 : \theta = b$ versus $K_1 : \theta \neq b$, we have* $P_{\{K_0:\theta=b\}}\big(2\min\{p_s(C_{lo}), \ p_s(C_{up})\} \leq \alpha\big) = \alpha$ and $2\min\{p_s(C_{lo}), p_s(C_{up})\} = 2\min\{H_n(b), 1 - H_n(b)\}$ is the *p-value of the corresponding test.*

A limiting result of the same nature as Theorem 2 holds for some more general null hypotheses. For instance, we can extend the above result to the case when K_0 is the union of finitely many disjoint closed intervals (bounded or unbounded) $C = \bigcup_{j=1}^k I_j$ with I_j being disjoint intervals of the type $(-\infty, a]$ or $[c, d]$ or $[b, \infty)$, or to the case when K_0 is $C = \{\theta_1, \theta_2, \ldots, \theta_k\}$. In the first case of C being a union of intervals (i.e., an intersection union test, see, Berger, 1982), both the rejection regions $\{\mathbf{x} : p_s(C) = \sum_{j=1}^k p_s(I_j) \leq \alpha\}$ and $\{\mathbf{x} : \max_{1\leq j\leq k} p_s(I_j) \leq \alpha\}$ correspond to level α tests asymptotically, under Condition (A$'$); Among the two, the test based on $\max_{1\leq j\leq k} p_s(I_j)$ is more powerful. In the second case of C being a collection of disjoint singletons, the rejection region $\left\{\mathbf{x} : \max_{1\leq j\leq k}\left[2\min\{p_s(C_{lo}^{(j)}), p_s(C_{up}^{(j)})\}\right] \leq \alpha\right\}$ corresponds to a level α test asymptotically. Here, $C_{lo}^{(j)} = (-\infty, \theta_j]$ and $C_{up}^{(j)} = [\theta_j, \infty)$, for $1 \leq j \leq k$. The claims are based on the following theorem.

THEOREM 3. *Suppose that Condition (A$'$) holds uniformly in θ lying in a compact subset of the parameter space. (i) For the test $K_0 : \theta \in C = \bigcup_{j=1}^k I_j$ versus $K_1 : \theta \in C^c$, where I_j are disjoint intervals of the type* $(-\infty, a]$ or $[c, d]$ or $[b, \infty)$, *we have* $\sup_{\theta\in C} P_\theta\big(p_s(C) \leq \alpha\big) \to \alpha$ and $\sup_{\theta\in C} P_\theta\big(\max_j p_s(I_j) \leq \alpha\big) \to \alpha$, as $n \to \infty$. *(ii) For the test $K_0 : \theta \in C = \{\theta_1, \ldots, \theta_k\}$*

versus $K_1 : \theta \in C^c$, *we have* $\max_{\theta \in C} P_\theta\big(\max_j \big[2\min\{p_s(C_{lo}^{(j)}),\, p_s(C_{up}^{(j)})\}\big] \le \alpha\big) \to \alpha$, as $n \to \infty$.

The proof of Theorem 3 is fairly straightforward, noting that under Condition (A$'$), the support of any set on the parameter space, say B, vanishes, unless the set B contains the true parameter value. This essentially reduces the problem to the cases discussed in Theorem 2. A formal proof can be found in Singh *et al.* (2007).

From the theorem, we can conclude that $\sum_{j=1}^{k} p_s(I_j) = \sum_{j=1}^{k} H_n(I_j)$ and $\max_{1\le j\le k} p_s(I_j) = \max_{1\le j\le k} H_n(I_j)$ are asymptotically equivalent p-values for tests involving a union of disjoint intervals, although the latter p-value is always numerically smaller than the prior one. Also, $\max_{1\le j\le k} [2\min\{p_s(C_{lo}^{(j)}),\, p_s(C_{up}^{(j)})\}] = \max_{1\le j\le k} \big[2\min\{H_n(\theta_j),\, 1 - H_n(\theta_j)\}\big]$ is an asymptotic p-value for the test involving a collection of singletons. Such type of tests have applications in bioequivalence problems and others.

5 Optimality (Comparison) of Confidence Distributions

As in the case of point estimation, we can have multiple confidence distributions for the same parameter under any specific setting. It begs a comparison of these different confidence distributions, in terms of their performance in statistical inference. Generally speaking, the more concentrated a confidence distribution is near the true parameter value θ_0, the more informative it is. In particular, given two confidence distributions for θ, say $H_1(\cdot)$ and $H_2(\cdot)$, one should compare $H_1(\theta_0 - \epsilon)$ to $H_2(\theta_0 - \epsilon)$ and also $1 - H_1(\theta_0 + \epsilon)$ to $1 - H_2(\theta_0 + \epsilon)$. In each case, the smaller the better, which typically translates into smaller confidence intervals and more powerful tests at the same level (or size). This reasoning motivates the following definition (see Schweder & Hjort (2002) and Singh *et al*. (2001, 2007)).

Definition 3: Given two confidence distributions $H_1(\cdot)$ and $H_2(\cdot)$ for θ, we say $H_1(\cdot)$ is superior to $H_2(\cdot)$ at $\theta = \theta_0$, if for all $\epsilon > 0$, $H_1(\theta_0 - \epsilon) \overset{sto}{\le} H_2(\theta_0 - \epsilon)$ and $1 - H_1(\theta_0 + \epsilon) \overset{sto}{\le} 1 - H_2(\theta_0 + \epsilon)$. Here, the symbol $\overset{sto}{\le}$ stands for stochastically less than or equal to.

Now, suppose $H_1(b)$ and $H_2(b)$ are simply p-values for the test $K_0 : \theta \le b$ versus $K_1 : \theta > b$ for any given b. Let $H_1(\cdot)$-based tests be more powerful than $H_2(\cdot)$-based tests, for any given b and at any level of significance. Moreover, assume a similar comparison holds for $1 - H_1(b)$ over $1 - H_2(b)$ for the test $K_0 : \theta \ge b$ versus $K_1 : \theta < b$. Then, it is a straightforward exercise to establish that $H_1(\cdot)$ is superior to $H_2(\cdot)$. This result leads to the following conclusion: The confidence distribution arising from the p-value function of a one-sided UMP (uniformly most powerful) family of tests is optimal in the sense that it is superior to any other exact confidence distribution for θ. A detailed derivation of this optimality via a sufficient statistics having MLR (monotone likelihood ratio property) can be found in Schweder & Hjort (2002) under the title Neyman-Pearson Lemma for confidence intervals.

To understand how a UMPU (uniformly most powerful unbiased) test leads to an optimal confidence distribution in the above sense, we note that, given a confidence distribution, $H(\cdot)$, the one-sided tests using $H(b)$ and $1 - H(b)$ as p-values for hypothesis tests $K_0 : \theta \le b$ versus $K_1 : \theta > b$ and $K_0 : \theta \ge b$ versus $K_1 : \theta < b$, respectively, are unbiased. The reason is that, for the first set of hypotheses where $H(b)$ is used as the p-values, the power of the test at a value in the alternative set, say $\theta = b + \epsilon$, is given by

$$P_{b+\epsilon}(H(b) < \alpha) \ge P_{b+\epsilon}(H(b+\epsilon) < \alpha) = \alpha.$$

The last equation is by the definition of a confidence distribution and α is the level of significance. A similar proof is given for the second set of hypotheses where $1 - H(b)$ is used as the p-value. This observation entails an immediate extension of the above optimality claim on the UMP tests. Suppose there is a family of one-sided tests which are UMPU leading to a confidence distribution $H^*(\cdot)$ as its p-value function. Given any other confidence distribution $H(\cdot)$, to be compared with $H^*(\cdot)$, we bring in focus the one-sided power unbiased tests based on $H(\cdot)$. Then, $H^*(\cdot)$ is superior to $H(\cdot)$ in terms of the power dominance due to the UMPU property of $H^*(\cdot)$. The conclusion is that a confidence distribution derived from a p-value function of a one-sided UMPU family of tests is optimal. See Singh *et al.* (2007) for additional details. Clearly, the UMPU class of tests is much larger than the UMP class of tests. For the normal sample case with known variance, the normal-based confidence distribution (due to its UMP property) is optimal. In case when σ^2 is unknown, the student t-based confidence distribution is optimal in view of the fact that the one-sided t-tests are UMPU (see Lehmann, 1991). Also, the χ^2-based CD for σ^2 is optimal since the one-sided χ^2 tests possess UMPU property.

In the context of optimality of confidence distributions for exponential families with nuisance parameters, Schweder & Hjort (2002) emphasized the importance of conditional tests. Suppose that the data $\mathbf{x}$ has a Lebesgue density of the form $f(\mathbf{x}) = \exp\left\{\theta s(\mathbf{x}) + \psi_1 A_1(\mathbf{x}) + \ldots + \psi_p A_p(\mathbf{x}) + B(\theta, \psi_1, \ldots \psi_p)\right\}$ for any given function B, where θ is scaler parameter of interest and $\Psi = \{\psi_1, \ldots, \psi_p\}$ are nuisance parameters. Then, the test with p-value

$$P_{(\theta,\Psi)}(s(\mathbf{x}) > s(\mathbf{x}_{obs}) \mid A_i(x) = A_i(\mathbf{x}_{obs}), i = 1, \ldots, p)$$

is exact and UMP. This yields a valuable tool for constructing optimal confidence distributions. See example 8 of Schweder & Hjort (2002) about bivariate Poisson sample with mean vector $(\lambda_i, \theta\lambda_i)$ where θ is the parameter of interest and λ_i's are nuisance parameters.

For confidence distributions, there are natural analogues of Pitman-type local and Bahadur-type non-local comparisons of tests (see, e.g., Rao, 1973). In particular, for the Pitman type local comparison between confidence distributions $H_1(\cdot)$ and $H_2(\cdot)$, one makes limiting stochastic comparison between $H_1(\theta_0 - \epsilon/\sqrt{n})$ and $H_2(\theta_0 - \epsilon/\sqrt{n})$ and also between $1 - H_1(\theta_0 + \epsilon/\sqrt{n})$ and $1 - H_2(\theta_0 + \epsilon/\sqrt{n})$. For Bahadur type non-local comparison, one compares almost-sure limits of $n^{-1}\log\{H_1(\theta_0 - \epsilon)\}$ and $n^{-1}\log\{H_2(\theta_0 - \epsilon)\}$ for a fixed $\epsilon > 0$. Similar comparisons are done between the upper tails of $H_1(\cdot)$ and $H_2(\cdot)$. These types of comparisons which clearly involve large deviation probabilities are recommended for exact confidence distributions only. See Singh *et al.* (2007) for more details.

There is an easy-to-deduce implication of the foregoing definition of $H_1(\cdot)$ being superior to $H_2(\cdot)$, in the sense of exact inference.

PROPOSITION 1. *Suppose that $H_1(\cdot)$ is superior to $H_2(\cdot)$ as defined in Definition 3, and both are strictly increasing. Then, for allt in* $(0, 1)$*, we have* $[H_1^{-1}(t) - \theta_0]^+ \overset{sto}{\leq} [H_2^{-1}(t) - \theta_0]^+$ *and* $[H_1^{-1}(t) - \theta_0]^- \overset{sto}{\leq} [H_2^{-1}(t) - \theta_0]^-$. *Thus,* $|H_1^{-1}(t) - \theta_0| \overset{sto}{\leq} |H_2^{-1}(t) - \theta_0|$.

The expression $|H^{-1}(t) - \theta_0|$ may be called deficiency of the one-sided confidence intervals $(-\infty, H^{-1}(t)]$ or $[H^{-1}(t), \infty)$. The logic behind this terminology is as follows. Consider a lower side confidence interval $(-\infty, H^{-1}(t)]$. If the interval covers θ_0, $[H^{-1}(t) - \theta_0]^+$ is the over-reach of the interval; If the interval does not cover θ_0, $[H^{-1}(t) - \theta_0]^-$ is the length by which it misses. In either case, a smaller $|H^{-1}(t) - \theta_0|$ is desirable. Thus, the smaller the deficiency is, the better the one-sided confidence interval is, given that the coverage requirement is met. The above proposition comes close to making a connection between a superior confidence distribution and a more concise confidence interval. Does a superior confidence distribution as defined in Definition 5.1 always provide two sided intervals with smaller (expected) length? The answer is still unclear to us without any additional conditions.

There is a natural notion of mean square error for a confidence distribution $H(\cdot)$:

$$\mathrm{MSE}(H) = \mathrm{E}_{\mathbf{x}} \int_{t \in \Theta} (t - \theta_0)^2 dH(t) = \mathrm{E}_{\xi}(\xi - \theta_0)^2,$$

where θ_0 is the true value of the parameter and ξ is the CD-random variable associated with $H(\cdot)$. It is an elementary exercise to show that, if confidence distribution $H_1(\cdot)$ is superior to another confidence distribution $H_2(\cdot)$ for the same parameter, then

$$\mathrm{MSE}(H_1) \leq \mathrm{MSE}(H_2).$$

This leads to a notion of optimality in terms of their mean squared errors.

6 Combining Confidence Distributions from Independent Sources

There are a number of developments and applications that highlight the potential and added values of confidence distributions as an effective tool for statistical inference. The following two sections are reviews of such developments, starting in this section with developments on combining confidence distributions. It is natural to combine confidence distributions from independent studies, and an approach that combines confidence distributions can potentially preserve more information than a traditional approach that combines just point estimators.

6.1 Combination of Confidence Distributions and a Unified Framework for Meta-Analysis

Suppose there are k independent studies that are dedicated to estimate a common parameter of interest θ. From the sample $\mathbf{x}_i$ of the ith study, we assume that we have obtained a confidence distribution $H_i(\cdot) = H_i(\mathbf{x}_i, \cdot)$ for θ. By extending the classical methods of combining p-values and based on a coordinate-wise monotonic function from $[0, 1]^k$ to $\mathbb{R} = (-\infty, +\infty)$, Singh *et al.* (2005) proposed a general recipe for combining these k independent confidence distributions:

$$H^{(c)}(\theta) = G_c\{g_c(H_1(\theta), \ldots, H_k(\theta))\}. \tag{7}$$

Here, $g_c(u_1, \ldots, u_k)$ is a given continuous function on $[0, 1]^k \to \mathbb{R}$ which is non-decreasing in each coordinate, and the function G_c is completely determined by the monotonic g_c function with $G_c(t) = P(g_c(U_1, \ldots, U_k) \leq t)$, where $U_1, \ldots, U_k$ are independent $U[0, 1]$ random variables. The function $H^{(c)}(\cdot)$ contains information from all k samples and it is referred to as a combined confidence distribution. When the underlying true parameter values, say $\theta_i^{(0)}$ for the ith study, of the k individual confidence distribution $H_i(\cdot)$ are the same (i.e., $\theta_i^{(0)} \equiv \theta_0$), it is evident that $H^{(c)}(\cdot)$ is a confidence distribution for the parameter θ (with true value θ_0). A nice feature of the combining method (7) is that it does not require any information regarding how the input confidence distributions $H_i(\cdot)$ are obtained, aside from the assumed independence. So it also provides a sensible way to combine input confidence distributions obtained from different procedures, even across different schools of inference.

Singh *et al.* (2005) studied Bahadur efficiency of the combination (7) in the setting where the underlying true parameter values of the k individual confidence distributions $H_i(\cdot)$ are the same. They showed that, when $g_c(u_1, \ldots, u_k) = F_{de}^{-1}(u_1) + \cdots + F_{de}^{-1}(u_k)$, the combined confidence distribution $H^{(c)}(\cdot)$ is most efficient, in terms of having the highest Bahadur slope (i.e., steepest exponential decay at the tails), among all possible combinations. The result is an extension of the theoretical result by Littell & Folks (1973) on the classical Fisher's p-value combination method. Here, $F_{de}(\cdot)$ is the cumulative distribution function of the standard double exponential (DE) distribution and k is fixed. Singh *et al.* (2005) also studied an alternative combination method of multiplying confidence densities, similar to likelihood combination of multiplying independent likelihood (density) functions. Singh *et al.* (2005) proved that a combined confidence distribution by the multiplication method is an exact confidence distribution in the cases of location parameters and scale parameters. They also obtained a related asymptotic result for a broader class of parameters. See also Lindley (1958) and Welch & Peers (1963) in the context of location and scale models.

Xie *et al.* (2011) further studied methods of combining confidence distributions under meta-analysis settings in which the underlying true parameters of the corresponding studies may or may not be the same (i.e., $\theta_i^{(0)} \not\equiv \theta_0$). Also, since Bahadur optimality may encounter technical problems for non-exact inferences (as is commonly the case in meta-analysis applications), Xie *et al.* (2011) concentrated on Fisher-type optimality (i.e., in terms of comparing exact/asymptotic variances of estimators), which is more relevant for confidence intervals, and provided a unifying framework for various existing meta-analysis approaches. Weighted combining to improve Fisher-efficiency of a combination plays a key role in this development, especially in model-based meta-analysis approaches.

Example 3 earlier illustrated that a p-value (significance) function is usually a confidence distribution or an asymptotic confidence distribution. Based on this observation, Xie *et al.* (2011) verified that all five classical approaches of combining p-values listed in Marden (1991), i.e., Fisher, Stouffer (normal), Tippett (min), Max, and Sum methods, can be subsumed under the framework of combining confidence distributions. Specifically, for Fisher, Stouffer (normal), Tippett (min), Max, and Sum p-value combinations, the g_c choices in (7) are $g_c(u_1, \dots, u_k) = F_{ne}^{-1}(u_1) + \dots + F_{ne}^{-1}(u_k)$ or $\Phi^{-1}(u_1) + \dots + \Phi^{-1}(u_k)$ or $\min(u_1, \dots, u_k)$ or $\max(u_1, \dots, u_k)$ or $u_1 + \dots + u_k$, respectively. Here, $F_{ne}^{-1}(u) = \log(u)$ and $\Phi^{-1}(u)$ are the inverse of the cumulative distribution function of the negative exponential and the standard normal distribution, respectively.

Also, Normand (1999) and Sutton & Higgins (2008) provided excellent reviews of model-based meta-analysis, for both fixed-effects and random-effects models, in modern biostatistics applications. Xie *et al.* (2011) illustrated that all five model-based meta-analysis estimators listed in Table IV of Normand (1999), i.e., the MLE, Bayesian and Method of Moment estimators under a fixed-effect model and the REML and Bayesian (normal prior) estimators under a random effect-model, can be obtained under the framework of combining confidence distributions (7) using the weighted recipe,

$$g_c(u_1, \dots, u_k) = w_1\Phi^{-1}(u_1) + \cdots + w_k\Phi^{-1}(u_k), \tag{8}$$

with different choices of weights w_i's (which are related to the variance estimators of the point estimators—see Xie *et al.* (2011) for further details). Since an appropriate choice of weights can improve the combining efficiency in normal models, this explains why a model-based approach is typically more efficient than an approach (classical unweighted) of combining p-values, when the model assumption holds. This development allowed, for the first time, the seemingly unrelated two classes of meta-analysis methods to be studied in a unified framework.

The framework of combining confidence distributions provides a structure and opportunity to explore and discover new sensible meta-analysis approaches. For instance, Xie *et al.* (2011) also developed two complementary robust meta-analysis approaches, with supporting asymptotic theory, under the unifying framework. In one approach, the size of each study goes to infinity, and in the other, the number of studies tends to infinity. The first approach takes advantage of the adaptive (data-dependent) weights in a special weighted combining approach of (7) with $g_c(u_1, \ldots, u_k) = w_1^{(a)} F_0^{-1}(u_1) + \cdots + w_k^{(a)} F_0^{-1}(u_k)$, for any given continuous cumulative distribution function F_0, where the adaptive weight $w_i^{(a)}$ converges to 0, asymptotically, when the underlying true value $\theta_i^{(0)}$ of the ith study is different from the target parameter value θ_0. See, Xie *et al.* (2011) for such a choice of adaptive weights. In the second approach, the g_c function in (7) is $g_c(u_1, \ldots, u_k) = w_1 u_1 + \ldots + w_k u_k$, where, for simplicity, only non-adaptive weights similar to those used in (8) were considered, although the development can be extended to include other weights. Both these robust meta-analysis approaches have high breakdown points and are resistant to studies whose underlying (unknown) true parameter values differ from that of the parameter of interest. Xie *et al.* (2011) proved that, in the first setting, the robust meta-analysis estimator is asymptomatically equivalent to the theoretically most efficient point estimator in fixed-effects models, regardless of whether any studies are outlying or not. The second approach can attain $(3/\pi)^{1/2} \approx 97.7\%$ efficiency asymptotically in both fixed-effects and random-effects models when there are no outlying studies. The second approach also offers some protection against model misspecification, and it has an interesting connection to an M-estimation approach.

A publicly available R-package has been developed by Yang *et al.* (2012) to implement this unifying framework for meta-analysis. It has a single R function, with different options, to perform meta-analysis of combining p-values, the model-based methods, the robust methods, Mantel-Haenszel method, Peto method, the exact method by Tian *et al.* (2009) for rare event data, among others.

6.2 Incorporation of Expert Opinions in Clinical Trials

External information, such as expert opinions, can play an important role in the design, analysis, and interpretation of clinical trials. Seeking effective schemes for incorporating external information with the primary outcomes has drawn increasing interest in pharmaceutical applications and other areas. Since opinions are not actual observed data from the clinical trials, it is a common belief that regular frequentist approaches are not equipped to deal with such a problem. Research in this area has been dominated by Bayesian approaches, and almost all methods currently used for such practices are Bayesian. In the Bayesian paradigm, as illustrated in Spiegelhalter *et al.* (1994), a prior distribution is first formed to express the initial beliefs concerning the parameter of interest based on either some objective evidence or some subjective judgment or a combination of both. Subsequently, clinical trial evidence is summarized by a likelihood function, and a posterior distribution is then formed by combining the prior distribution with the likelihood function.

Indeed, incorporating evidence from external sources has long been considered the strong suit of Bayesian analysis, while being one of the weak points of frequentist theory. But the more recent developments on confidence distributions contain several attempts, namely Bickel (2006), Xie *et al.* (2013) and Singh & Xie (2012), to overcome this difficulty in the school of frequentist inference. These CD-based approaches can be outlined as follows: a confidence distribution is first used to summarize the prior information or expert opinions. Subsequently, clinical trial evidence is summarized by a confidence distribution, often based on a likelihood function. A combined confidence distribution is then obtained by combining these two confidence distributions. This combined confidence distribution can thereby be used to derive various inferences, and its role in the frequentist inference is similar to that of a posterior distribution in the Bayesian inference. These developments provide an example that confidence distributions can provide useful statistical inference tools for problems where frequentist methods were previously unavailable.

In particular, Bickel (2006) focused on the setting of a normal clinical trial, in which the author used an objective Bayesian argument to justify his treatment of the prior information from expert opinions as a confidence distribution and then used the DE-based recipe proposed in Singh *et al.* (2005) to combine it with a normal confidence distribution from the clinical trial. The research of Xie *et al.* (2013) stemmed from a consulting project on a real clinical trial of a migraine therapy in a pharmaceutical company, in which the information solicited from expert opinions needs to be incorporated with the information from a clinical study of a migraine therapy with binary outcomes. They illustrated, using solely frequentist modelling arguments, that information from experts (based on the standard Baysesian design of Parmar *et al.* (1994) and Spiegelhalter *et al.* (1994)) can be summarized as a confidence distribution. This confidence distribution is then combined with a confidence distribution from the data. Singh & Xie (2012) extended the concept of confidence distribution to a typical Bayesian setting where parameters are random, and used the recipe (7) to combine a prior distribution with a confidence distribution from data. The outcome function is called a CD posterior, an alternative to Bayes posterior, which was shown to have the same coverage property as the Bayes posterior, in terms of the joint distribution of the random parameter and data. The approaches in Bickel (2006) and Xie *et al.* (2013) can be viewed as a compromise between the Bayesian and frequentist paradigms. For instance, we consider the CD-based approach proposed in Xie *et al.* (2013) as a frequentist approach, since the parameter is treated as a fixed value and not a random entity. But, nonetheless, it also has a Bayesian flavor, since the prior expert opinions represent only the relative experience or prior knowledge of the experts but not any actual observed data.

The CD-based approaches are easy to implement and computationally cheap (no MCMC algorithm), even in non-normal case. Xie *et al.* (2013) and Singh & Xie (2012) highlighted the advantage of a CD-based approach that it does not require any prior on nuisance parameters, compared to a Bayesian approach. More importantly, in their binomial setting, Xie *et al.* (2013) discovered a counterintuitive "discrepant posterior phenomenon" that is inherent in a Bayesian approach but can be avoided in a CD-based approach. We describe below the discrepant posterior phenomenon using a simple binomial clinical trial.

Consider the scenario of a binomial clinical trial, in which we have x_0 successes from Binomial (n_0, p_0) in the control group and x_1 successes from Binomial (n_1, p_1) in the treatment group. In addition, we have an informative prior (expert opinions) on the parameter of interest $\delta = p_1 - p_0$, say $\pi(\delta)$, and also perhaps a marginal prior of p_0, say $\pi(p_0)$, from historical experiments, but we do not completely know the joint prior of (p_0, p_1). Clearly, we can not directly apply Bayes formula $f(\delta|data) \propto \pi(\delta) f(data|\delta)$ focusing only on δ, because it is not possible to find a conditional density function $f(data|\delta)$ in this trial. A full Bayesian solution is to jointly model both p_0 and p_1, as the model used in Joseph *et al.* (1997) with independent beta priors $\pi(p_0, p_1) = \pi(p_0)\pi(p_1)$ for p_0 and p_1; A constraint is that the marginal prior of δ from the assumed $\pi(p_0, p_1)$ needs to agree with the known $\pi(\delta)$. Based on this independent beta prior model, Xie *et al.* (2013) provided a numerical example, as illustrated in Figure 2 (a)–(b), where strangely the marginal posterior of δ is more extreme than both the prior and the profile likelihood of δ, so that an estimate from the posterior (e.g., by mode, $\hat{\delta}_{post} = 0.211$) is much bigger than those from the prior and the likelihood (e.g., by mode, $\hat{\delta}_{prior} = 0.128$ and $\hat{\delta}_{like} = 0.104$, respectively). That is, the experts suggest about 12.8% improvement and the clinical evidence suggests about 10.4% improvement but, putting them together, the posterior suggests that the overall estimator of the treatment effect is 21.1%, which is much bigger than either that reported by the experts or that suggested by the clinical data. This is certainly counterintuitive. Xie *et al.* (2013) provided a detailed discussion that this counterintuitive phenomenon is intrinsically

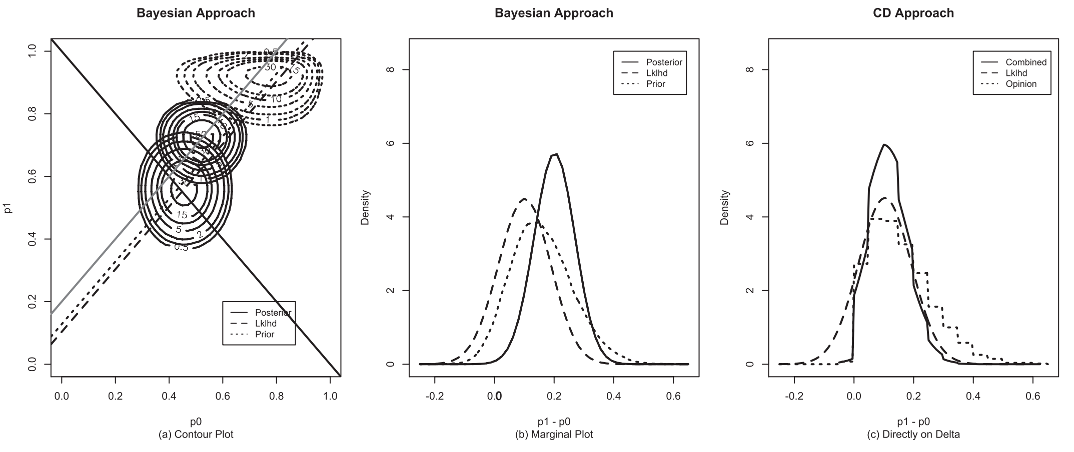

Figure 2. *(a) Contour plots of the joint prior $\pi(p_0, p_1)$, density function $f(data|p_0, p_1)$, and posterior $f(p_0, p_1|data)$; (b) Projections (marginals) of $\pi(p_0, p_1)$, $f(data|p_0, p_1)$, and $f(p_0, p_1|data)$* onto the direction of $\delta = p_1 - p_0$*; both (a) and (b) are from a numerical example with data* $(x_0, x_1, n_0, n_1) = (31, 33, 68, 59)$ *and an independent beta prior with* BETA(14.66, 4.88) *for* $\pi(p_0)$ and BETA(46.81, 4.68) for $\pi(p_1)$*; (c) a reproduction of Figure 1(e) of Xie et al. (2013), where a survey of expert opinions on δ (summarized in a histogram) is directly combined with a confidence distribution for δ from a clinical data.*

mathematical and also general. They suggested that the phenomenon may occur in at least a certain direction as long as skewed distributions are involved in multivariate Bayesian analysis, regardless of which prior, model, or parameterization is used.

This discrepant posterior phenomenon posts a challenging question for Bayesian analysis in this particular application. Note that we only have informative marginal priors on δ and p_0 but not their joint prior. If we can come up with a joint prior for (p_0, p_1) or (p_0, δ) whose marginal priors match with $\pi(p_0)$ and $\pi(\delta)$ and also this joint prior is "truthful" (whatever this means), then the posterior outcome is trustworthy (whatever the outcome is and regardless whether the marginal posterior is outlying or not). In this case, some researchers may argue that the phenomenon shows the power of the joint Bayesian method in extracting information that otherwise may be overlooked, even if it is counterintuitive. However, the two marginal priors on p_0 and $\delta = p_1 - p_0$ can not fully determine the joint prior. How can we tell whether we have got a "truthful" joint prior (even if it has the same two marginal priors)? If we are not so sure about the joint prior, should we still trust the posterior obtained by using this joint prior, especially when we have a discrepant posterior? Alternatively, should we question about the joint prior when the discrepant posterior phenomenon occurs? In practice, we may be inclined to accept a posterior result that sits between the prior and the data evidence and we may also tend to think that there is a "truthful" joint prior that can yield this in-between outcome. It is unclear, outside the normal case, how we can practically come up with a joint prior that is in agreement with the likelihood evidence so that the discrepant posterior phenomenon does not occur.

As a frequentist procedure, the CD-based method can bypass this difficult task of jointly modelling (p_0, p_1) or (p_0, δ) and focus directly on the parameter of interest δ (c.f., the discussion of "division of labor" in the second half of Section 3.3). Figure 2 (c) illustrated a CD-based method (Xie *et al.*, 2013) for the same problem, where a survey of expert opinions on δ (summarized in a histogram) is directly combined with a confidence distribution for δ from a clinical data. Depending on the perceived "truthful" model of the joint distribution of (p_0, p_1), it is still debatable in this particular example whether the marginal posterior (as that in

Figure 2 (b)) or the combined confidence distribution (as that in Figure 2 (c)) should be used. But the CD-based method certainly enjoys the benefit of "division of labor" and is free of the counterintuitive discrepant posterior phenomenon.

7 CD-Based New Methodologies, Examples, and Applications

This section reviews several general inferential tools and new methodologies, utilizing confidence distributions. Additional examples and developments of confidence distributions in various applications are also provided.

7.1 CD-Based Likelihood Calculations

We have discussed in Section 2 that, given a likelihood function, we may normalize it to obtain an (often asymptotic) confidence distribution. In this subsection, we present that, conversely, given a confidence distribution, we could also try to construct certain types of likelihood functions for inference. Two notable examples are Efron (1993), which used a confidence distribution to derive a so-called implied likelihood function, and Schweder & Hjort (2002), which proposed a notion of reduced likelihood function based on the development of confidence distributions.

The objective of Efron (1993) was to produce an approximate likelihood function for a scalar parameter of interest, say, θ in a multiparameter family, with all nuisance parameters eliminated. To achieve this goal, the starting point is a confidence distribution from "any system of approximate confidence intervals for the single parameter θ". In particular, let $h_{\mathbf{x}}(\theta)$ be the density function of a confidence distribution for the parameter of interest θ. Efron (1993) suggested to modify this confidence density by introducing a doubled data set $(\mathbf{x}, \mathbf{x})$ via a "data-doubling device". This data-doubling device is "similar to that used in bias-reduction techniques such as Richardson extrapolation and the jackknife", in which "we imagine having observed two independent data sets, both of which happen to equal the actual data set $\mathbf{x}$". From the doubled data set $(\mathbf{x}, \mathbf{x})$, we get another confidence density function for θ, $h_{\mathbf{xx}}(\theta)$, using the same system of confidence intervals. The implied likelihood function is then defined by

$$L^{\dagger}_{\mathbf{x}}(\theta) = \frac{h_{\mathbf{xx}}(\theta)}{h_{\mathbf{x}}(\theta)}.$$

To illustrate the procedure, Efron (1993) provided a simple example for a parameter $\theta =$ Prob(head), the probability of landing on head, on the basis of observing x heads out of n independent flips of a coin. Based on the estimator $\hat{\theta} = x/n$, we can obtain an asymptotic confidence distribution with density $h_{x,n}(\theta) \propto \theta^{x}(1-\theta)^{n-x}\left\{(1-\hat{\theta})/(1-\theta) - \hat{\theta}/\theta\right\}$ and, from the doubled data set of $2n$ coin flips with $2x$ observed heads, another confidence distribution with density $h_{2x,2n}(\theta) \propto \theta^{2x}(1-\theta)^{2(n-x)}\left\{(1-\hat{\theta})/(1-\theta) - \hat{\theta}/\theta\right\}$. Thus, the implied likelihood is

$$L^{\dagger}_{x,n}(\theta) = \frac{h_{2x,2n}}{h_{x,n}(\theta)} \propto \theta^{x}(1-\theta)^{n-x}.$$

Efron (1993) noted that this implied likelihood is equal to the actual likelihood function for θ in this binomial example, but the confidence distribution based on $\hat{\theta} = x/n$, i.e., $h_{x,n}(\theta)$, is not.

Efron (1993) suggested using the implied likelihood function as a tool for making "Bayesian, empirical Bayesian and likelihood inferences about θ", where the likelihood formula has all nuisance parameters eliminated. A Bayesian analysis based on this likelihood "requires a prior distribution only for θ, and not for the entire unknown parameter vector". The method is mostly effective for multiparameter exponential families, where there exists a simple and accurate system of approximate confidence distributions for any smoothly defined parameter. Efron (1993) illustrated his development using the so called ABC intervals (c.f., Efron, 1987; DiCiccio & Efron, 1992). He also explored a Bayesian motivation for the development, and also justified the implied likelihood in terms of high-order adjusted likelihoods and the Jeffreys-Welch and Peers theory of uninformative priors (Welch & Peers, 1963). It was proved that the ABC implied likelihood agrees with the profile likelihood of θ up to the second order in the repeated sampling situation in a multiparameter exponential family (Efron, 1993, theorem 2). Interestingly, this connection to the profile likelihood function also implies that the implied likelihood function is also an asymptotic confidence distribution itself, after a normalization with respect to the parameter θ. Thus, we consider the implied likelihood as a modified confidence distribution from the original confidence distribution $h_{x,n}(\theta)$. The advantage to use an implied likelihood is that there is no need to involve nuisance parameters.

Schweder & Hjort (2002) proposed a so-called reduced likelihood function that was determined via a pivot from a scalar statistic, along with a confidence distribution. Schweder & Hjort (2002) highlighted that the reduced likelihood is a "proper" likelihood and it is free of all nuisance parameters. Together with the corresponding confidence density, it forms "a very useful summary for any parameter of interest". Specifically, let $\psi(T_n, \theta)$ be a pivot or an approximate pivot statistic with a continuous cumulative distribution function F_n and density function f_n, where T_n is a one-dimensional statistic. Also, assume $\psi(T_n, \theta)$ is monotonic (say, increasing) in θ. The probability density function of T_n is $f(t; \theta) = f_n\big(\psi(t, \theta)\big)\big|\frac{\partial}{\partial t}\psi(t, \theta)\big|$, which, when viewed as a function of θ, is called a reduced likelihood function by Schweder & Hjort (2002). Note that $H_{piv}(\theta) = F_n\big(\psi(T_n, \theta)\big)$ is the confidence distribution induced by $\psi(T_n, \theta)$, and this reduced likelihood function can be re-expressed as

$$L_{t_n}(\theta) = f_n\Big(F_n^{-1}(H_{piv}(\theta))\Big)\Big|\frac{\partial}{\partial t}\psi(t_n, \theta)\Big| = h_{piv}(\theta)\Big|\frac{\partial}{\partial t}\psi(t, \theta)\Big|\Big/\Big|\frac{\partial}{\partial \theta}\psi(t, \theta)\Big|\Big|_{t=t_n}, \quad (9)$$

where $h_{piv}(\theta)$ is the density function of $H_{piv}(\theta)$. Like Efron's implied likelihood, the reduced likelihood is free of nuisance parameters and it gives "a summary of the data that allows the future user to give the past data the appropriate weight in the combined analysis".

From (9), the reduced likelihood $L_{t_n}(\theta)$ and the confidence density $h_{piv}(\theta)$ differs by a factor of $c_{n,\theta} = \big|\frac{\partial}{\partial t}\psi(t, \theta)\big|\big/\big|\frac{\partial}{\partial \theta}\psi(t, \theta)\big|\big|_{t=t_n}$. Schweder & Hjort (2002) stated that, by reparametrization, the reduced likelihood $L_{t_n}(\theta)$ could be brought to be proportional to $h_{piv}(\theta)$, the confidence density. In particular, in the case when $c_{n,\theta}$ is increasing in the reparameterization $\mu = a(\theta)$ and $\frac{\partial}{\partial \theta}\mu \propto c_{n,\theta}$, the implied likelihood function for μ is same as the confidence density for μ, apart from a constant not involving the parameter, both of which are proportional to $f_n\big(\psi(T_n, a^{-1}(\mu)\big)$. Also, in the case when the parameter θ can be consistently estimated, we often have $c_{n,\theta} = c_n\{1 + o_p(1)\}$ for a constant c_n. In this case, the implied likelihood function is asymptotically proportional to the confidence density $h_{piv}(\theta)$, without any reparameterization. Again, we consider the reduced likelihood as a modified confidence distribution from the original $H_{piv}(\theta)$, which is obtained directly from the pivot.

Efron (1993) demonstrated a use of the implied likelihood function in an empirical Bayes example, which incorporated an empirical Bayes prior from past clinical trials with an implied likelihood from a "current" study on a new treatment for recurrent bleeding ulcer using the Bayes formula. Singh *et al.* (2005) re-analyzed the same data under a purely frequentist setting using the CD combining recipe discussed in Section 5.1 (with and without (adaptive) weights). Numerical contrasts with Efron's empirical Bayes approach was also provided. Schweder (2003) applied the methods that were developed in Schweder & Hjort (2002) for reduced likelihood and confidence distributions to provide model estimates for abundance of a closed stratified population of bowhead whales off Alaska.

7.2 Confidence Curve

The confidence curve was first proposed by Birnbaum (1961) as an estimate of an "omnibus form", which "incorporates confidence limits and intervals at various levels and a median-unbiased point estimate, together with critical levels of tests of hypothesis of interest, and representations of the power of such tests". Although the original proposal of the confidence curve did not use the notion of a confidence distribution, the concept arise out of the same consideration to incorporate confidence intervals of all levels. It can be easily defined using a confidence distribution function. Let $H_n(\cdot)$ be a confidence distribution for parameter θ. The confidence curve is just then

$$CV_n(\theta) = 1 - 2|H_n(\theta) - 0.5| = 2\min\{H_n(\theta), 1 - H_n(\theta)\}.$$

If $H_n(\cdot)$ is a p-value function from an one-sided hypothesis test, $CV_n(\cdot)$ then corresponds to the p-values of two-sided tests for singleton hypotheses; see, e.g., Fraser (1991) and also Section 3 of this article. As previously stated, a confidence distribution may be loosely viewed as a "piled-up" form of one-sided confidence intervals of all levels. In the same spirit, the confidence curve may be viewed as a "piled-up" form of two-sided confidence intervals of equal tails at all levels.

Figure 3(a) plots, in a solid curve, a confidence distribution for the θ in Example 7, and the solid curve in Figure 3(b) is the corresponding confidence curve. In Figure 3(a), drawing two lines at the height of $\alpha/2$ and $1 - \alpha/2$, respectively, produces two intersections with the confidence distribution. The two points on the x-axis corresponding to the intersections are the lower and upper bounds of the level $1 - \alpha$ equal tailed two-sided confidence interval. In Figure 3(b), drawing a single line across the height of α in Figure 3(b) provides two intersections with the confidence curve. The two points on the x-axis corresponding to two intersections are the two end points of the same level $1 - \alpha$ confidence interval of θ. When the height α increases to one, the confidence curve in Figure 3(b) reduces to a focal point, the median of the confidence distribution $M_n = H_n^{-1}(1/2)$, which is a median unbiased estimator of θ. Schweder *et al.* (2007) described the confidence curve as a "funnel plot pointing at the median of the confidence distribution". Blaker & Spjøtvoll (2000) argued that a single point estimate or a confidence interval or a testing conclusion may be too crude to be a summary of the information given by the data. They suggested plotting a confidence curve to show "how each parameter value is ranked in view of the data". They also provided examples whereby using a confidence curve could "alleviate" a problem in cases in which confidence sets are empty or include every parameter value. See, also, a tutorial by Bender *et al.* (2005) on using confidence curves in medical research, in which the authors demonstrated that confidence curves are useful complements to the conventional methods of presentation.

A notable application of confidence curve is due to Blaker (2000), in which the concept of confidence curve is instrumental in the methodological development of finding improved exact confidence intervals for a general discrete distribution. Blaker (2000) noted that the standard exact confidence intervals tend to be very conservative and too wide with coverage probabilities strictly larger than the designed nominal level, especially in small and moderate sample sizes. He utilized confidence curves to obtain improved exact confidence intervals for discrete distributions. The approach can be illustrated using a binomial example with sample x, sample size n and success probability p. The p-value function from the one-sided test $K_0 : p \le b$ versus $K_1 : p > b$ is an asymptotic confidence distribution $H_{(x,n)}(b) = H_n(x, b) = 1 - \sum_{k=0}^{x-1} \binom{n}{k} b^k (1-b)^{n-k}$

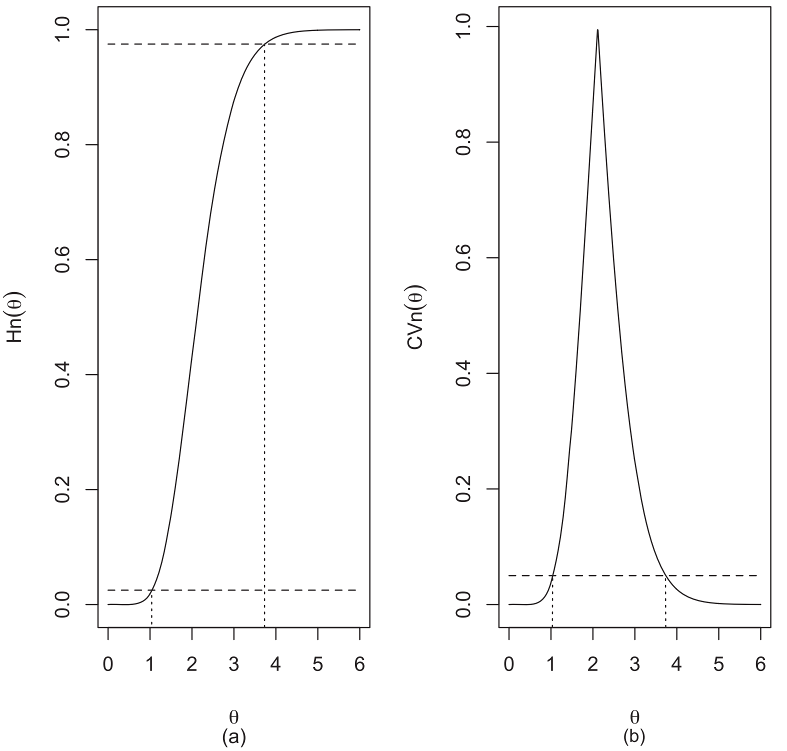

Figure 3. *(a) A confidence distribution and (b) a confidence curve for the scale parameter of a gamma distribution based on a sample of size $n = 5$ from Gamma$(2, \theta)$; see also Example 7.*

and the confidence curve is $CV(b) = 2\min\left\{H_{(x,n)}(b), 1 - H_{(x,n)}(b)\right\}$. However, confidence intervals based on $CV(b)$ only have correct asymptotic coverage probabilities. To guarantee strictly the coverage probability, one often has to enlarge the confidence intervals, which can be deduced by modifying the confidence curve $CV(b)$ to one of its asymptotically equivalent copies

$$CV^{\dagger}(b) = \begin{cases} 2H_{(x,n)}(b) & \text{if } H_{(x,n)}(b) < 1 - H_{(x+1,n)}(b), \\ 2\{1 - H_{(x+1,n)}(b)\} & \text{if } H_{(x,n)}(b) > 1 - H_{(x+1,n)}(b), \\ 1 & \text{if } H_{(x,n)}(b) = 1 - H_{(x+1,n)}(b). \end{cases} \tag{10}$$

See, e.g., Blaker (2000). The level $1 - \alpha$ equal tailed two-sided confidence interval from $CV^{\dagger}(b)$ is $\mathcal{A} = \left\{p : CV^{\dagger}(p) \geq \alpha\right\} = \left\{p :\ H^{-1}_{(x,n)}(\alpha/2) \leq p \leq H^{-1}_{(x+1,n)}(1 - \alpha/2)\right\}$, which is the well-known Clopper-Pearson exact confidence interval. Blaker (2000) noted that the equal-tailed requirement is too restrictive, thus considered a further modified confidence curve function

$$CV^{\ddagger}(b) = \begin{cases} H_{(x,n)}(b) + H_{(\zeta_1,n)}(b) & \text{if } H_{(x,n)}(b) < 1 - H_{(x+1,n)}(b), \\ \left\{1 - H_{(x+1,n)}(b)\right\} + \left\{1 - H_{(\zeta_2,n)}(b)\right\} & \text{if } H_{(x,n)}(b) > 1 - H_{(x+1,n)}(b), \\ 1 & \text{if } H_{(x,n)}(b) = 1 - H_{(x+1,n)}(b), \end{cases}$$

where $\zeta_1 = \sup\{u : P_{\{p=t\}}(X \geq x) > P_{\{p=t\}}(X \leq u)\} = \sup\{u : H_{(x,n)}(p) \geq 1 - H_{(u,n)}(p)\}$ and $\zeta_2 = \sup\{u : P_{\{p=t\}}(X \geq x + 1) > P_{\{p=t\}}(X \leq u)\} = \sup\{u : 1 - H_{(x+1,n)}(p) \geq H_{(u,n)}(p)\}$.

Blaker (2000) proved that intervals obtained from $CV^{\ddagger}(b)$ can guarantee the exact coverage probabilities and also $CV^{\ddagger}(b) \leq CV^{\dagger}(b)$. Thus, while maintaining the right coverage, the class of exact confidence intervals from $CV^{\ddagger}(p)$ is shorter than the Clopper-Pearson exact confidence intervals. Blaker (2000) also showed that this class of exact intervals satisfies the nesting condition, which is violated by several other exact intervals, including those by Sterne (1954), Crow (1956), Blyth & Still (1983), and Casella (1986). Here, the nesting condition refers to the requirement that intervals with a larger confidence level always include those with a smaller confidence level.

Schweder (2007) called the function $CN(\theta) = 1 - CV(\theta) = 1 - 2\min\left\{H_n(\theta), 1 - H_n(\theta)\right\}$ a confidence net and extended it to a multidimensional case with θ being a $p \times 1$ vector. He suggested obtaining a confidence net by mapping the profile deviance $D(\theta) = -2\{\ell_n(\theta) - \ell_n(\hat{\theta})\}$ into interval (0, 1) by its cumulative distribution function, and used the ABC-method and bootstrapping (c.f., Efron, 1987; DiCiccio & Efron, 1992; Schweder & Hjort, 2002) to illustrate how to obtain a family of simultaneous confidence bands for multiple parameters. Here, $\ell_n(\theta)$ is the log profile likelihood function and $\hat{\theta}$ is the MLE of θ. Confidence net for a multiparameter vector is related to the version of a circular confidence distribution (c-CD) proposed by Singh *et al.* (2001, 2007) using the concept of data depth. In particular, a sample-dependent function $H_n(\cdot) = H_n(\mathbf{x}, \cdot)$ on $\Theta \in \mathbb{R}^p$ is called a *confidence distribution in the circular sense* (c-CD) for $p \times 1$ multiparameter θ, if (i) it is a probability distribution function on the parameter space Θ for each fixed sample set $\mathbf{x}$, and (ii) the $100(1-\beta)\%$ central region of $H(\theta)$ is a confidence region for θ, having the coverage level $100(1-\beta)\%$ for each β. Here, the central region refers to a centrality function defined through a given measure of data depth: $C(\theta) = P\{\eta : D(\eta) \leq D(\theta)\}$, where both the data depth D and the probability P are computed based on H_n (as a distribution function on $\Theta \in \mathbb{R}^p$). See Liu *et al.* (1999) among others, for various concepts of data-depth. In a one-dimensional (scalar) situation, the centrality function $C(\theta)$ coincides with the confidence curve $C(\theta) = CV(\theta) = 2\min\{H_n(\theta), 1 - H_n(\theta)\}$. This type of multiparameter confidence distribution only covers a family of center outwards regions (not all Borel type of regions) in the multidimensional case.

7.3 CD-Based Simulation Methods

As described in Section 3.1, given a confidence distribution, a CD-random variable can typically be generated, and this CD-random variable is closely related to and has similar theoretical properties as a bootstrap estimator. Simulating CD-random variables offers a new model-based Monte Carlo approach that is similar to the bootstrap, but without directly involving data. Since a CD-random variable is not limited to being just a bootstrap estimator, this Monte Carlo approach can be more flexible and go beyond the standard bootstrap procedures.

An example of such a development is Claggett *et al.* (2012), who developed a CD-based Monte Carlo method to make inference for extrema of parameters, such as $\max\{\theta_1, \ldots, \theta_k\}$, and more generally for any order statistics of parameters $\theta_{(m)}$ = the mth smallest among $\{\theta_1, \ldots, \theta_k\}$. Here, $\theta_1, \ldots, \theta_k$ are unknown underlying parameters of k independent studies. Hall & Miller (2010) mentioned that the problem of making inference for extrema of parameters is one of "the existing problems where standard bootstrap estimators are not consistent and where alternative approaches also face significant challenges". To overcome this problem associated with the standard bootstrap methods, Claggett *et al.* (2012) considered a more flexible CD-based Monte Carlo method. Unlike the standard bootstrap methods that directly use the empirical distribution of the bootstrap estimator of $\theta_{(m)}$, say $\theta^*_{(m)}$, to make inference, Claggett *et al.* (2012) proposed using the empirical distribution of a weighted sum of CD-random variables $\xi^* = \sum_{i=1}^k w_i \xi_i \big/ \sum_{i=1}^k w_i$, where the CD random variables ξ_i's are simulated from given confidence distributions for the respective studies. Noting that $\xi_{(m)}$ = the mth smallest of $\{\xi_1, \ldots, \xi_k\}$ behaves the same as the bootstrap estimator $\theta^*_{(m)}$, this CD-based method is equivalent to the standard bootstrap method if the weights are chosen as $w_i = \mathbf{1}_{\{\xi_i = \xi_{(m)}\}}$. But, with many possibilities for choosing weights, the CD-based method is more flexible. Under the standard asymptotically normal case and with a suitable set of weight choices, Claggett *et al.* (2012) were able to prove that ξ^* provides an asymptotically valid inference for $\theta_{(m)}$ in a large sample setting. (In the setting, the empirical weights asymptotically go to 0 when θ_i is outside the tie or near tie set; see Claggett *et al.*, 2012 and also Xie *et al.*, 2009). They also explored the theoretical performance of the standard bootstrap method in the presence of ties or near ties among the θ_i's, and showed that the weight corresponding to the standard bootstrap method (i.e., $w_i = \mathbf{1}_{\{\xi_i = \xi_{(m)}\}}$) does not satisfy the conditions required to make a correct inference for $\theta_{(m)}$.

The highly cited Monte Carlo method by Parzen *et al.* (1994) is also in essence a method of simulating CD-random variables (but in multidimensional case), which are in turn used to provide inference (variance estimates) for regression parameters in quantile and rank regression models. Their simulation method starts by assuming an exact or asymptotic pivotal estimating function, say $S_{\mathbf{x}}(\theta)$, which at the true parameter value $\theta = \theta_0$ can be generated by a random vector $\mathbf{z}$ whose distribution function is completely known or can be estimated consistently. Parzen *et al.* (1994) proposed simulating a large number of independent copies of $\mathbf{z}$ and solving the equation $S_{\mathbf{x}}(\theta) = \mathbf{z}$ for θ^*_z. These copies of θ^*_z are then used to make inference, e.g., use the sample variance of θ^*_z to estimate the variance of $\hat\theta$. Here, $\hat\theta$ is the point estimator that solves $S_{\mathbf{x}}(\theta) = 0$. In their proof, Parzen *et al.* (1994) required $n^{1/2}B(\hat\theta - \theta) = -S_{\mathbf{x}}(\theta) + o(1 + ||n^{1/2}(\hat\theta - \theta)||_2)$, where B is a positive definite constant matrix. To understand the essential connection of this approach to the CD-based Monte Carlo method without being bogged down by technical details associated with the multidimensional case, let us assume for simplicity that θ, $\mathbf{z}$ and B are scalars. Also, denote by $G(\cdot)$ the cumulative distribution function of $\mathbf{z}$. In this case, $G\big(n^{1/2}B(\theta_0 - \hat\theta)\big) \approx G\big(S_{\mathbf{x}}(\theta_0)\big)$ and the latter is distributed as $G(\mathbf{z}) \sim U[0, 1]$. Thus, $H_n(\theta) = G\big(n^{1/2}B(\theta - \hat\theta)\big)$ is an asymptotic confidence distribution for the parameter θ. The copies of θ^*_z simulated by Parzen *et al.* (1994) are nothing but realizations of copies of a CD-random variable from $H_n(\theta)$. This simulation has a fiducial flavor as the authors acknowledged in their original paper, although the approach does not need any fiducial reasoning in either their Monte Carlo simulation or their mathematical proofs.

The class of CD-based Monte Carlo approaches can be considered an extension of the well-studied and widely applied bootstrap methods. As a new simulation mechanism, considerably more constructive development is still needed. But we believe that the flexibility (not tied to the interpretation as bootstrap estimators) and the dependence only on data summary (not the original data) of this class of approaches will prove to be a useful simulation tool in a variety of applications. Also, see Kim & Lindsay (2010), who used (simulated) realizations of asymptotically normally distributed CD-random vectors to facilitate a visual analysis of confidence sets generated by an inference function such as a likelihood or a score function.

7.4 Additional Examples and Applications of CD-Developments

There are several other developments of confidence distributions targeting specific areas of applications. For instance, in the field of finance and economics, Coudin & Dufour (2004) used confidence distributions to facilitate a nice development of robust estimation of regression parameters in a linear median regression model. They relied on an empirical p-value function from a sign-based test for a two-sided hypothesis to make inference on the parameters of interest. This empirical p-value function is related in essence the confidence curve function described in 7.2, but in a multiparameter setting. In their development, no assumption was made on the shape of the error distribution, allowing for heterogeneity (or heteroskedasticity) of unknown form, non-continuous distributions, and very general serial dependence (linear and nonlinear). The empirical p-value function was derived from Monte Carlo sign-based tests, which have correct size in finite samples under the assumption that the disturbances have a null conditional median. Consistency, asymptotic normality and a connection to Hodges-Lehman-type estimators were provided. The authors applied the methods successfully to two studies of empirical financial and macroeconomic data: a trend model for the S&P index, and an equation used to study β-convergence of output levels across the states in the US.

Besides the papers already mentioned in Sections 6 and 7, there are also several other publications developing and applying confidence distributions in the fields related to medical and health research. For example, under the setting of clinical trials, Mau (1988) used a confidence distribution to measure the strength of evidence for practically equivalent treatments in terms of efficacy. The article described the concept of confidence distribution with clarity and also touched the topic of combining confidence distributions in the setting under consideration. Tian *et al.* (2011) used a multivariate CD concept to obtain optimal confidence regions for a vector of constrained parameters in a setting of survival studies. They showed in an analysis of a survival data set that the volume of the resulting 95% confidence region is only one-thirty-fourth of that of the conventional confidence region. Their confidence region also has better frequency coverage than the corresponding Bayesian credible region.

Developments and applications of confidence distributions can also be found in the fields of agriculture, fisheries research, and others. Schweder *et al.* (2010) proposed a set of complex capture-recapture models and used confidence distributions to make inferences on abundance, mortality, and population growth of bowhead whales. The authors analyzed 10 years of systematic photographic surveys conducted during the spring migrations of bowhead when whales pass Point Barrow, Alaska. Efron's BC and ABC methods (c.f., Efron, 1987 and Schweder & Hjort, 2002) were used to construct confidence distributions and confidence curve for several parameters of interest. Their inferences supported the view that photo surveys at Barrow are valuable for estimating abundance and demographic parameters for the population of bowhead whales in Bering-Chukchi-Beaufort seas. In addition, see Bityukov *et al.* (2007) and also Bityukov *et al.* (2010) for an encounter of confidence distributions and the use of Poisson example, with a masking background, in an application in the field of nuclear physics.

8 Summary

In this paper we have presented the concept of a confidence distribution, entirely within the frequentist school and logic, and reviewed recent developments together with their applications. We have shown how confidence distributions, as a broad concept, can subsume and be associated to many well-known notions in statistics across different schools of inference. We also showed how one could derive almost all types of frequentist inference from a confidence distribution and highlighted its promising potential, as an effective inferential tool. It is our belief that developments of confidence distributions can provide new insights towards many developments in both classical and modern statistical inferences.

As an emerging new field of research, there are many important topics and interesting questions yet to be addressed. For instance, as an estimation method conceptually not different from point estimation, many important questions and theoretical developments under the framework of decision theory are yet to be tackled. We reviewed a couple of results on optimality. But topics such as admissibility, minimax estimation, invariance, and prediction are yet untouched in the developments of confidence distributions. Additionally, as stated in Schweder & Hjort (2002), a simultaneous confidence distribution for multiple parameters can be "difficult" to define. It is still an open question how (or whether) one can define a multivariate confidence distribution, in a general non-Gaussian setting of exact inference, to ensure that their marginal distributions are confidence distributions for the corresponding single parameters. As an example, consider the standard Behrens-Fisher setting, where a joint confidence distribution (fiducial distribution) of the two population means (μ_1, μ_2) has a joint density of the form $f_1(\frac{\mu_1-\bar{x}_1}{s_1/\sqrt{n_1}})f_2(\frac{\mu_2-\bar{x}_2}{s_2/\sqrt{n_2}})/\{s_1 s_2/\sqrt{n_1 n_2}\}$. The marginal distribution of $\mu_1 - \mu_2$ is only an asymptotic confidence distribution (as both sample sizes $n_i \to \infty$) but not a confidence distribution in the exact sense. Here, $\bar{x}_i$ and s_i^2 are sample mean and variance for the respective study and f_i is the density function for the student's t-distribution with $n_i - 1$ degrees of freedom, $i = 1, 2$. The good news in the multidimensional case is that under asymptotic settings or wherever bootstrap theory applies, we can still work with multivariate confidence distributions with ease. In addition, if we limit ourselves to center-outwards confidence regions (instead of all Borel sets) in the $p \times 1$ parameter space, concepts such as the c-CDs considered in Singh *et al.* (2001, 2007) and the confidence net considered in Schweder (2007), offer coherent notions of multivariate confidence distributions in the exact sense.

In the practice of point estimation, a particular type of estimator, say for example an MLE or an unbiased estimator, may not exist in some special settings. Similarly, it is conceivable that we may encounter difficulty in obtaining a sensible confidence distribution under some settings. Cox (1958) commented on "whether confidence interval estimation is at all satisfactory" in "complicated cases, where the upper 5 per cent limit" is "larger than the upper 1 per cent limit". In this situation when confidence intervals violate the nesting condition, we may not be able to directly translate the set of intervals into a confidence distribution. But we may resort to an alternative approach to find confidence intervals that satisfy the nesting condition thus obtain a confidence distribution. (For instance, in the binomial example consider by Blaker (2000) in Section 7.2, the classical tests such as those by Sterne (1953) and others violate the nesting condition, and it is hard to directly relate them to a confidence distribution. But we can alternatively use the Blaker method to get nested confidence intervals and thus a confidence distribution easily.) An alternative approach is to mold the concept of confidence distribution to "upper" and "lower" versions. We may study the latter in a future project.

Finally, the discussion in this article assumes that the underlying model is more or less given, and we address the inferential problem given that the assumed model is correct. Theoretical frameworks for development and applications of confidence distributions are not yet developed to address the very important questions of model uncertainty, model diagnosis and model selection. We look forward to seeing developments along these lines which, we believe, can bear rich fruits for modern statistical applications.

Acknowledgements

This research is partly supported by research grants from NSF (DMS1107012, DMS0915139, SES0851521) and NSA (H98230-08-1-0104). The authors thank Professor Marc Hallin, the AE, and two reviewers for their constructive suggestions. They also thank Professors David R. Cox, Tore Schweder, Nils Hjort, Bill Strawderman, Regina Liu, and David Hoaglin for their discussions, comments, and suggestions.

References

Babu, G.J. & Singh, K. (1983). Inference on means using the bootstrap. *Ann. Statist.*, **11**, 999–1003.

Barndorff-Nielsen, O.E. & Cox, D.R. (1994). *Inference and Asymptotics*. London: Chapman & Hall.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.*, **53**, 370–418; **54**, 296–325. Reprinted in *Biometrika*, **45** (1958), 293–315.

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

Berger, R.L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics*, **24**, 295–300.

Bickel, D.R. (2006). Incorporation expert knowledge into frequentist inference by combining generalized confidence distributions. Available at: http://128.84.158.119/abs/math/0602377v1.

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

Bityukov, S., Krasnikov, N., Nadarajah, N. & Smirnova, V. (2010). Confidence distributions in statistical inference. In *Proceedings of the 30th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Eds. A. Mohammad-Djafari , J.-F. Bercher & P. Bessiere, pp. 446–456. Melville, NY.

Bityukov, S.I., Krasnikov, N.V., Smirnova, V.V. & Taperechkina, V.A. (2007). The transform between the space of observed values and the space of possible values of the parameter. *Proc. Sci. (ACAT)*, **62**, 1–9.

Blaker, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *Canad. J. Statist.*, **28**, 783–798.

Blaker, H. & Spjøtvoll, E. (2000). Paradoxes and improvements in interval estimation. *Amer. Statist.*, **54**, 242-247.

Blyth, C.R. & Still, H. (1983). Binomial confidence intervals. *J. Amer. Statist. Assoc.*, **78**, 108-116.

Casella, G. (1986). Refining binomial confidence intervals. *Canad. J. Statist.*, **14**, 113–129.

Clagget, B., Xie, M. & Tian, L. (2012) . Nonparametric inference for meta analysis with fixed unknown, study-specific parameters. Harvard University Biostatistics Working Paper Series. Working Paper 154. Available at http://biostats.bepress.com/harvardbiostat/paper154/.

Coudin, E. & Dufour, J.M. (2004). Hodges-Lehmann sign-based estimators and generalized confidence distributions in linear median regressions with heterogeneous dependent errors. CIREQ/CIRANO, Department of Economics, McGill University, 49 pp. (revised 2010).

Cox, D.R. (1958). Some problems with statistical inference. *Ann. Math. Stat*., **29**, 357–372.

Cox, D.R. (2006). *Principles of Statistical Inference.* London: Cambridge University Press.

Cox, D.R. & Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman & Hall.

Crow, E.L. (1956). Confidence intervals for a proportion. *Biometrika*, **43**, 423–435.

David, H.A. & Edwards, A.W.F. (2001). *Annotated Readings in the History of Statistics*. New York: Springer-Verlag.

Dempster, A.P. (2008). The Dempster-Shafer calculus for statisticians. *Internat. J. Approx. Reason.*, **48**, 365–377.

DiCiccio, T. & Efron, B. (1992). More accurate confidence intervals in exponential families. *Biometrika*, **79**, 231–45.

Efron, B. (1986). Why isn't everyone a Bayesian? *Amer. Statist.*, **40**, 262–266.

Efron, B. (1987). Better bootstrap confidence intervals and bootstrap approximations. *J. Amer. Statist. Assoc.,* **82**, 171–185.

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

Efron, B. (1998). R.A.Fisher in the 21st century. *Stat. Sci.*, **13**, 95–122.

Efron, B. & Tibshirani, R.J. (1994). *An Introduction to the Bootstrap.* London: Chapman & Hall.

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

Fisher, R.A. (1930). Inverse probability. *Proc. Cambridge Philos. Soc.*, **26**, 528–535.

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

Fisher, R.A. (1960). On some extensions of Bayesian inference proposed by Mr. Lindley. *J. R. Stat. Soc.* B, **22**, 299–301.

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

Fraser, D.A.S. (1991). Statistical inference: Likelihood to significance. *J. Amer. Statist. Assoc.*, **86**, 258–265.

Fraser, D.A.S. (2011). Is Bayes posterior just quick and dirty confidence? *Stat. Sci.*, **26**, 299–316.

Fraser, D.A.S. & Mcdunnough, P. (1984). Further remarks on asymptotic normality of likelihood and conditional analyses. *Canad. J. Statist.*, **12**, 183–190.

Hall, P. (1992). On the removal of skewness by transformation. *J. R. Stat. Soc. Ser.* B, **54**, 221–228.

Hall, P. & Miller, H. (2010). Bootstrap confidence intervals and hypothesis tests for extrema of parameters. *Biometrika*, **97**, 881–892.

Hampel, F. (2006). The proper fiducial argument. *Inf. Transf. Combin.*, LNCS 4123, pp. 512–526.

Hannig, J. (2009). On generalized fiducial inference. *Statist. Sinica*, **19**, 491–544.

Johnson, R.A. (1967). An asymptotic expansion for posterior distributions. *Ann. Math. Stat.*, **38**, 1899–1906.

Joseph, L., du Berger, R. & Belisle, P. (1997). Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Stat. Med.*, **16**, 769–781.

Kass, R. (2011). Statistical inference: The big picture. *Statist. Sci.*, **26**, 1–9.

Kendall, M. & Stuart, A. (1974). *The Advanced Theory of Statistics*, 2, 3rd ed. London: Griffin.

Kim, D. & Lindsay, B.G. (2011). Using confidence distribution sampling to visualize confidence sets. *Statist. Sinica*, **21**, 923–948.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications in Statistics*, **1**, 277–330.

LeCam, L. (1958). Les propriétés asymptotiques des solutions de Bayes. *Plubl. Inst. Statist. Univ. Paris*, **7**, 17–35.

Lehmann, E.L. (1991). *Testing Statistical Hypotheses.* New York: Springer-Verlag.

Littell, R.C. and Folks, J.L. (1973). Asymptotic optimality of Fisher's method of combining independent tests. II. *J. Amer. Statist. Assoc.*, **68**, 193–194.

Lindley, D.V. (1958). Fiducial distribution and Bayes theorem. *J. R. Stat. Soc. Ser.* B, **20**, 102–107.

Liu, R.Y., Parelius, J. & Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.,* **27**, 783–858.

Mau, J. (1988). A statistical assessment of clinical equivalence. *Stat. Med.*, **7**, 1267–1277.

Marden, J.I. (1991). Sensitive and sturdy p-values. *Ann. Statist*. **19**, 918–934.

Martin, R., Zhang, J. & Liu, C. (2010). Dempster-Shafer theory and statistical inference with weak beliefs. *Statist. Sci.*, **25**, 72–87.

Neyman, J. (1934). On the two different aspects of representative method: The method of stratified sampling and the method of purpose selection. *J. Roy. Statist. Soc. Ser. A,* **97**, 558–625.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philos.Trans. R. Soc.* A., **237**, 333–380.

Normand, S.-L. (1999). Meta-analysis: Formulating, evaluating, combining, and reporting. *Stat. Med.*, **18**, 321–359.

Parmar, M.K.B., Spiegelhalter, D.J, Freedman, L.S. & Chart Steering Committee (1994). The chart trials: Bayesian design and monitoring in practice. *Stat. Med.*, **13**, 1297–1312.

Parzen, M., Wei, L.J. & Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika*, **81**, 341–350.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley & Sons.

Reid, N. & Fraser, D.A.S. (2010). Mean likelihood and higher order approximations. *Biometrika*, **97**, 159–170.

Schweder, T. (2003). Abundance estimation from multiple photo surveys: Confidence distributions and reduced likelihoods for bowhead whales off Alaska. *Biometrics*, **59**, 974–983.

Schweder, T. (2007). Confidence nets for curves. *Advances in Statistical Modeling and Inference. Essays in honor of Kjell A. Doksum.* World Scientific. 593–609.

Schweder, T. & Hjort, N.L. (2002). Confidence and likelihood. *Scand. J. Stat.*, **29**, 309–332.

Schweder, T & Hjort, N.L (2003). Frequentist analogues of priors and posteriors. In *Econometrics and the Philosophy of Economics*, Ed. B.P. Stigum, pp. 285–317. Princeton, New Jersey: Princeton University Press.

Schweder, T., Sadykova, D., Rugh, D. & Koski, W. (2010). Population estimates from aerial photographic surveys of naturally and variably marked bowhead whales. *J. Agric. Biol. Environ. Stat.*, **15**, 1–19.

Singh, K. & Xie, M. (2011). Discussions on professor Fraserâs article on "Is Bayes posterior just quick and dirty confidence?" *Stat. Sci*., **26**, 319–321.

Singh, K. & Xie, M. (2012). CD-posterior — combining prior and data through confidence distributions. In *Contemporary Developments in Bayesian Analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman.* IMS Collection, Eds. D. Fourdrinier, E. Marchand & A.L. Rukhin, Vol. 8, pp. 200–214. Baltimore: Mattson Publishing Services.

Singh, K., Xie, M. & Strawderman, W. (2001). Confidence distributions–concept, theory and applications. Technical Report, Dept. of Statistics, Rutgers University. Updated 2004.

Singh, K., Xie, M. & Strawderman, W.E. (2005). Combining information from independent sources through confidence distributions. *Ann. Statist.* **33**, 159–183.

Singh, K., Xie, M. & Strawderman, W.E. (2007). Confidence distribution (CD)-distribution estimator of a parameter. In *Complex Datasets and Inverse Problems*. IMS Lecture Notes-Monograph Series, 54, 132–150.

Spiegelhalter, D.J., Freedman, L.S. & Parmar, M.K.B. (1994). Bayesian approaches to randomized trials. *J. Roy. Statist. Soc. Ser. A*, **157**, 357–416.

Sterne, T.H. (1954). Some remarks on confidence or fiducial limits. *Biometrika*, **41**, 275–278.

Sutton, A.J. & Higgins, J.P.T. (2008). Recent developments in meta-analysis. *Stat. Med.*, **27**, 625–650.

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

Tian, L., Wang, R., Cai, T. & Wei, L.J. (2011). The highest confidence density region and its usage for joint inferences about constrained parameters. *Biometrics*, **67**, 604–610.

Wasserman, L. (2007). Why isn't everyone a Bayesian? In *The Science of Bradley Efron*, Eds. C.N. Morris & R. Tibshirani, pp. 260–261. New York: Springer.

Welch, B.L. & Peers, H.W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. R. Stat. Soc. Ser. B*, **25**, 318–29.

Wong, A.C.M. (1993). A note on inference for the mean parameter of the gamma distribution. *Stat. Probab. Lett.*, **17**, 61–66.

Wong, A.C.M. (1995). On approximate inference for the two-parameter gamma model. *Stat. Papers*, **36**, 49–59.

Xie, M., Liu, R.Y., Damaraju, C.V. & Olson, W.H. (2013). Incorporating external information in analyses of clinical trials with binary outcomes. Available at http://www.imstat.org/aoas/next_issue.html.

Xie, M., Singh, K. & Strawderman, W.E. (2011). Confidence distributions and a unifying framework for meta-analysis. *J. Amer. Statist. Assoc.*, **106**, 320–333.

Xie, M., Singh, K. & Zhang, C.-H. (2009). Confidence intervals for population ranks in the presence of ties or near ties. *J. Amer. Statist. Assoc.*, **104**, 775–788.

Yang, G., Shi, P. & Xie, M. (2012). *gmeta*: Meta-Analysis via a Unified Framework under Confidence Distribution (R package). Version 2. Available at http://stat.rutgers.edu/home/mxie/packages/gmetaRpackage/.

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

Résumé

Il est courant, en inférence fréquentielle, d'utiliser un point unique (une estimation ponctuelle) ou un intervalle (intervalle de confiance) dans le but d'estimer un paramètre d'intérêt. Une question très simple se pose: peut-on également utiliser, dans le même but, et dans la même optique fréquentielle, à la façon dont les Bayésiens utilisent une loi a posteriori, une distribution de probabilité? La réponse est affirmative, et les distributions de confiance apparaissent comme un choix naturel dans ce contexte. Le concept de distribution de confiance a une longue histoire, longtemps associée, à tort, aux théories d'inférence fiducielle, ce qui a compromis son développement dans l'optique fréquentielle. Les distributions de confiance ont récemment attiré un regain d'intérêt, et plusieurs résultats ont mis en évidence leur potentiel considérable en tant qu'outil inférentiel. Cet article présente une définition moderne du concept, et examine les ses évolutions récentes. Il aborde les méthodes d'inférence, les problèmes d'optimalité, et les applications. A la lumière de ces nouveaux développements, le concept de distribution de confiance englobe et unifie un large éventail de cas particuliers, depuis les exemples paramétriques réguliers (distributions fiducielles), les lois de rééchantillonnage, les *p*-valeurs et les fonctions de vraisemblance normalisées jusqu'aux a priori et posteriori bayésiens. La discussion est entièrement menée d'un point de vue fréquentiel, et met l'accent sur les applications dans lesquelles les solutions fréquentielles sont inexistantes ou d'une application difficile. Bien que nous attirions également l'attention sur les similitudes et les différences que présentent les approches fréquentielle, fiducielle, et Bayésienne, notre intention n'est pas de rouvrir un débat philosophique qui dure depuis près de deux cents ans. Nous espérons bien au contraire contribuer à combler le fossé qui existe entre les différents points de vue.

[*Received August 2011, accepted November 2012*]