Confidence Distribution and Distribution Estimation for Modern Statistical Inference

29

Yifan Cui and Min-ge Xie

Contents

Abstract

This chapter introduces to readers the new concept and methodology of confidence distribution and the modern-day distributional inference in statistics. This discussion should be of interest to people who would like to go into the depth of the statistical inference methodology and to utilize distribution estimators in practice. We also include in the discussion the topic of generalized fiducial inference, a special type of modern distributional inference, and relate it to the concept of confidence distribution. Several real data examples are also provided for practitioners. We hope that the selected content covers the greater part of the developments on this subject.

Y. Cui (✉)
Zhejiang University, Hangzhou, Zhejiang, China
e-mail: cuiyf@zju.edu.cn

M. Xie
Rutgers University, New Brunswick, NJ, USA
e-mail: mxie@stat.rutgers.edu

29.1 Introduction

A confidence distribution (CD) refers to a sample-dependent distribution function that can represent confidence intervals (regions) of all levels for a parameter of interest [74, 90]. Instead of the usual point estimator or confidence interval, CD is a distribution estimator of a parameter of interest with a pure frequentist interpretation. The development of the CD can be traced back to, for example, [16, 28, 47, 66]. However, its associated inference schemes and applications have not received much attention until the recent surge of interest in the research of CD and its applications [25, 46, 52, 53, 72–74, 77, 78, 82, 90, 91, 94]. All of these developments of CDs, along with a modern definition and interpretation, provide a powerful inferential tool for statistical inference.

One of the main contributions of CD is its applications on fusion learning [12, 15, 40, 51–53, 72, 75, 77, 81, 91, 92]. Combining CDs from independent studies naturally preserves more information from the individual studies than a traditional approach of combining only point estimators. A unified framework of combining CDs for fusion learning generally includes three steps: (1) using a CD to summarize relevant information or obtain an inference result from each study, (2) combining information from different sources or studies by combining these CDs, and (3) making inference via the combined CD. This approach has sound theoretical support and has been applied to many practical situations with much success.

On a different note, the fiducial distribution may be considered as one special type of CD, which provides a

systematic way to obtain a CD. The origin of fiducial inference can be traced back to R.A. Fisher [28] who introduced the concept of a fiducial distribution for one parameter and proposed the use of this fiducial distribution to avoid the problems related to the choice of a prior distribution. Since the mid-2000s, there has been a renewed interest in modifications of fiducial inference [2, 7, 8, 22, 24, 30, 31, 33–37, 41, 56, 57, 59–63, 68, 73, 74, 79, 85, 87, 90, 93, 96].

We briefly overview these modern approaches which extend Fisher's original fiducial argument. We then focus on a recent development termed generalized fiducial inference and its applications [14, 17, 37, 41, 42, 44, 49, 50, 65, 86, 88, 89] that greatly expand the applicability of fiducial ideas. We demonstrate this recipe on several examples of varying complexity. The statistical procedures derived by the generalized fiducial inference often have very good performance from both theoretical and numerical points of view.

29.2 Confidence Distribution

29.2.1 The Concept of CD

This section will mainly focus on the concept of CD. The CD can be viewed as a distribution estimator, which can be utilized for constructing statistical procedures such as point estimates, confidence intervals, hypothesis tests, etc. The basic notion of CDs is related to the fiducial distribution of [28]; however, it is a pure frequentist concept. Some have suggested to view CD as the frequentist analog of Bayesian posterior distribution [e.g., 73, 74]. More broadly, if the credible intervals or regions obtained from a Bayesian posterior match with frequentist intervals or regions (either exactly or asymptotically), then the Bayesian posterior can be viewed as CD, and thus Bayesian approach is also a way to obtain CD [90].

Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed and $\mathcal{X}$ is the sample space corresponding to the dataset $(X_1, X_2, \ldots, X_n)$. Let θ be a scalar parameter of interest and Θ be the parameter space. The following formal definitions of CD and asymptotic CD are proposed in [72,77].

Definition 29.2.1 (CD and Asymptotic CD) A function $H_n(\cdot) = H_n(x, \cdot)$ on $\mathcal{X} \times \Theta \to [0, 1]$ is called a CD for a parameter, if (1) for each given $x \in \mathcal{X}$, $H_n(\cdot)$ is a (continuous) cumulative distribution function on Θ and (2) at the true parameter value $\theta = \theta_0$, $H_n(\theta_0) \equiv H_n(x, \theta_0)$, as a function of the sample x, follows the uniform distribution $U(0, 1)$. In addition, the function $H_n(\cdot)$ is called an asymptotic CD if condition (2) is replaced by (2') at the true parameter $\theta = \theta_0$, $H_n(\theta_0) \xrightarrow{d} U(0, 1)$ as $n \to \infty$.

From a nontechnical point of view, a CD is a function of both the parameter and the sample which satisfies two conditions. The first condition basically states that for any fixed sample, a CD is a distribution function on the parameter space. The second condition essentially requires that the corresponding inference derived by a CD has desired frequentist properties. Section 29.2.2 will further discuss how to use the second condition to extract information from a CD to make inference.

Birnbaum [9] introduced the concept of confidence curve as "an omnibus technique for estimation and testing statistical hypotheses," which was independent of the development of CD. From a CD $H_n(\theta)$, the confidence curve can be written as

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

Indeed, confidence curve is an alternative expression of CD and it is a very useful graphical tool for visualizing CDs. On a plot of $CV_n(\theta)$ versus θ, a line across the y-axis of the significance level α, for any $0 < \alpha < 1$, intersects with the confidence curve at two points, and these two points correspond to an $1 - \alpha$ level, equal-tailed, two-sided confidence interval for θ. In addition, the maximum of a confidence curve is the median of the CD which is the recommended point estimator.

We present below five illustrating examples of CDs. More examples refer to [74, 77, 90].

Example 29.2.1 Suppose the data $X_i \sim N(\mu, 1), i = 1, \ldots, n$, with unknown μ. Let $\bar{x}_n$ denote the sample mean. Then $N(\bar{x}_n, 1/n)$ is a CD for μ, and it can be represented in the following three forms: (i) confidence distribution (cumulate distribution form), $H_n(\mu) = \Phi(\sqrt{n}(\mu - \bar{x}_n))$; (ii) confidence density (density form), $h_n(\mu) = \frac{1}{\sqrt{2\pi/n}} \exp\{-\frac{n}{2}(\mu - \bar{x}_n)^2\}$; and (iii) confidence curve, $CV_n(\mu) = 2\min\{\Phi(\sqrt{n}(\mu - \bar{x}_n)), 1 - \Phi(\sqrt{n}(\mu - \bar{x}_n))\}$. See Fig. 29.1 for an illustration. The data are generated from $N(0.3, 1)$ with sample size 100.

Example 29.2.2 ([77]) Suppose the data $X_i \sim N(\mu, \sigma^2), i = 1, \ldots, n$, with both unknown μ and σ. A CD for μ is $H_n(\mu) = F_{t_{n-1}}(\frac{\sqrt{n}(\mu - \bar{x}_n)}{s_n})$, where s_n is the sample standard deviation and $F_{t_{n-1}}(\cdot)$ is the cumulative distribution function of student t distribution with parameter $n - 1$. A CD for σ^2 is $H_n(\sigma^2) = 1 - F_{\chi^2_{n-1}}(\frac{(n-1)s_n^2}{\sigma^2})$, where $F_{\chi^2_{n-1}}(\cdot)$ is the cumulative distribution function of the χ^2_{n-1}-distribution.

Example 29.2.3 ([77]) Let $\widehat{\theta}$ be a consistent estimator of θ. For bootstrap, the distribution of $\widehat{\theta}^* - \theta$ is estimated by the bootstrap distribution $\widehat{\theta}^* - \widehat{\theta}$, where $\widehat{\theta}^*$ is the estimator of θ computed on a bootstrap sample [26]. An asymptotic CD for θ is given by $H_n(\theta) = 1 - \Pr(\widehat{\theta}^* - \widehat{\theta} \leq \widehat{\theta} - \theta) =$
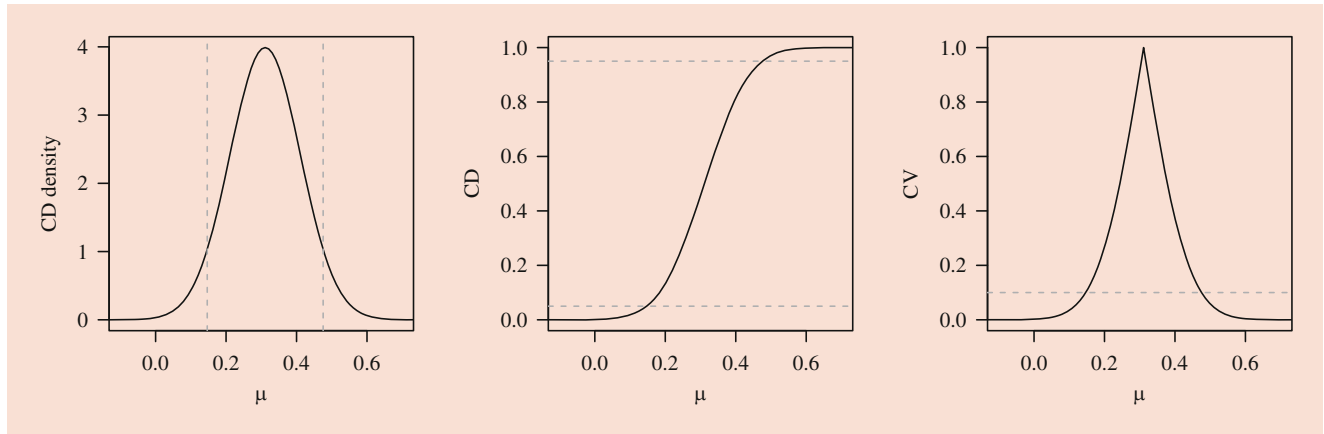
Fig. 29.1 Confidence distribution presented in Example 29.2.1 in the forms of density function, cumulative distribution function, and confidence curve

$\Pr(\widehat{\theta}^* \geq 2\widehat{\theta} - \theta)$. In addition, when the limiting distribution of normalized $\widehat{\theta}$ is symmetric, the raw bootstrap distribution $H_n(\theta) = 1 - \Pr(\widehat{\theta} - \widehat{\theta}^* \leq \widehat{\theta} - \theta) = \Pr(\widehat{\theta}^* \leq \theta)$ is also an asymptotic CD.

Example 29.2.4 Suppose we are interested in the location parameter θ of a continuous distribution. When the distribution F is symmetric, i.e., $F(\theta - y) = 1 - F(\theta + y)$, θ is the median. The Wilcoxon rank test for $H_0 : \theta = t, H_1 : \theta \neq t$ is based on the summation of signed ranks of $Y_i - t$, i.e., the test statistic $W = \sum_{i=1}^n Z_i R_i$, where R_i is the rank of $|Y_i - t|$ and Z_i is an indicator variable with 1 if $Y_i - t > 0$ and -1 otherwise. Denote by $p(t)$ the p-value associated with the Wilcoxon rank test for $H_0 : \theta = t, H_1 : \theta \neq t$. When t varies in $(-\infty, \infty)$, the p-value $p(t)$ is referred to as a p-value function. We can prove that the p-value function $p(t)$ is an asymptotic CD [90]. Figure 29.2 provides illustrations of the asymptotic CD density $p'(t)$, the asymptotic CD function $p(t)$, and the asymptotic CV $2 \min\{p(t), 1 - p(t)\}$ for two sample sizes. The data are generated from $N(0, 1)$ with sample sizes $n = 10$ and 100, respectively.

Example 29.2.5 ([78]) Suppose that there is an independent and identically distributed sample of size n from a semi-parametric model involving multiple parameters. Let $l_n(\theta)$ be the log profile likelihood function and $\mathcal{J}_n(\theta) = -\ddot{l}_n(\theta)$ be the observed Fisher information for a scalar parameter of interest θ. Under certain mild assumptions, Theorem 4.1 of [78] proves that, for any given θ,

$$G_n(\theta) = H_n(\theta) + o_p(1), \text{ where } G_n(\theta) = \frac{\int_{-\infty}^\theta \exp\{l_n(x)\}dx}{\int_{-\infty}^\infty \exp\{l_n(x)\}dx},$$

$$H_n(\theta) = \Phi\left(\frac{\theta - \widehat{\theta}}{\sqrt{\mathcal{J}_n(\widehat{\theta})/n}}\right), \quad \widehat{\theta} = \arg\max_\theta l_n(\theta).$$

Because 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 $U(0, 1)$. Thus, $G_n(\theta)$ is an asymptotic CD. From this observation, we see that CD-based inference may subsume a likelihood inference in some occasions.

If the sample X is from a discrete distribution, we can typically invoke a large sample theory to obtain an asymptotic CD to ensure the asymptotic frequentist coverage property, when the sample size is large. However, when the sample size is limited, we sometimes may want to exam the difference between the "distribution estimator" and the $U(0, 1)$ distribution to get a sense of under and over coverage. To expand the concept of CD to cover the cases of discrete distributions with finite sample sizes, we introduce below the notions of lower and upper CDs. The lower and upper CDs provide us inference statements that are associated with under and over coverages at every significant level.

Definition 29.2.2 (Upper and Lower CDs) A function $H_n^+(\cdot) = H_n^+(x, \cdot)$ on $\mathcal{X} \times \Theta \to [0, 1]$ is said to be an upper CD for a parameter, if (i) for each given $x \in \mathcal{X}$, $H_n(\cdot)$ is a monotonic increasing function on Θ with values ranging within $(0, 1)$ and (ii) at the true parameter value $\theta = \theta_0$, $H_n^+(\theta_0) \equiv H_n^+(x, \theta_0)$, as a function of the sample x, is stochastically less than or equal to a uniformly distributed random variable $U \sim U(0, 1)$, i.e.,

$$\Pr\left(H_n^+(X, \theta_0) \leq t\right) \geq t. \tag{29.1}$$

Correspondingly, a lower CD $H_n^-(\cdot) = H_n(x, \cdot)$ for parameter θ can be defined but with (29.1) replaced by $\Pr\left(H_n^-(X, \theta_0) \leq t\right) \leq t$ for all $t \in (0, 1)$.

More generally, we also refer to $H_n^+(\cdot)$ and $H_n^-(\cdot)$ as the upper and lower CD, respectively, even when the monotonic
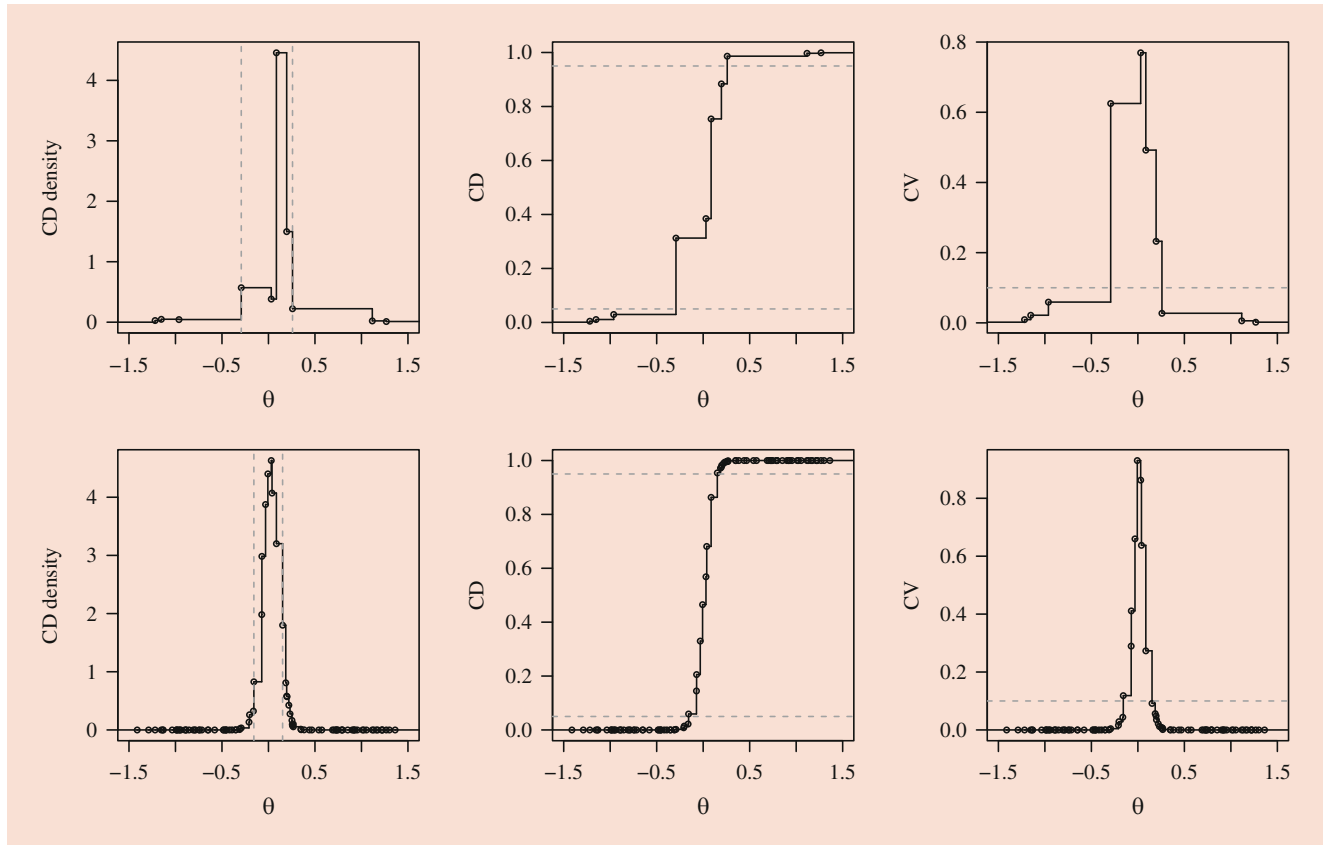
Fig. 29.2 Confidence distributions presented in Example 29.2.4 in the forms of density function, cumulative distribution function, and confidence curve. The top row is for sample size $n = 10$ and the bottom row is $n = 100$

condition (i) is removed. Note that, due to the stochastic dominance inequalities in the definition, we have, for any $\alpha \in (0, 1)$,

$$\Pr\left(\theta_0 \in \left\{\theta : H_n^+(X, \theta) \leq \alpha\right\}\right) \geq \alpha \text{ and}$$
$$\Pr\left(\theta_0 \in \left\{\theta : H_n^-(X, \theta) \leq \alpha\right\}\right) \leq \alpha.$$

Thus, a level-$(1 - \alpha)$ confident interval (or set) $\{\theta : H_n^+(X, \theta) \leq 1 - \alpha\}$ or $\{\theta : H_n^-(X, \theta) \geq \alpha\}$ has guaranteed the coverage rate of $(1 - \alpha)100\%$, regardless of whether we have the monotonic condition in (i). After we remove the monotonic condition in (i), $H_n^+(\cdot)$ and $H_n^-(\cdot)$ may not be a distribution function, and the "nest-ness property" of confidence intervals/sets may also be lost. Here, the "nest-ness property" refers to "a level-$(1 - \alpha)$ confidence set $C_{1-\alpha}$ is not necessarily inside its corresponding level-$(1 - \alpha')$ confidence set $C_{1-\alpha'}$, when $1 - \alpha < 1 - \alpha'$."

To conclude this section, we present an example of lower and upper CDs.

Example 29.2.6 ([40]) Suppose sample X is from Binomial (n, p_0) with observation x. Let $H_n(p, x) = \Pr(X > x) = \sum_{x<k\leq n} \binom{n}{k} p^k (1 - p)^{n-k}$. We can show that $P(H_n(p_0, X) \leq t) \geq t$ and $P(H_n(p_0, X - 1) \leq t) \leq t$. Thus, $H^+(p, x) =$

$H_n(p, x)$ and $H_n^-(p, x) = H_n(p, x - 1)$ are lower and upper CDs for the success rate p_0. The half-corrected CD [25,37,72] is

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

29.2.2 CD-Based Inference

Analogous to the Bayesian posterior, a CD contains a wealth of information for constructing any type of frequentist inference. We illustrate three aspects of making inference based on a given CD. Figure 29.3 from [90] provides a graphical illustration of the point estimation, confidence interval, and hypothesis testing. More specifically:

Point Estimation The natural choices of point estimators of the parameter θ given a CD $H_n(\cdot)$ include (i) the median $\widetilde{\theta}_n = H_n(1/2)$, (ii) the mean $\bar{\theta}_n = \int_{\theta\in\Theta} \theta \, dH_n(\theta)$, and (iii) the mode $\widehat{\theta}_n = \arg\max_{\theta\in\Theta} h_n(\theta)$, where $h_n(\theta) = dH_n(\theta)/d\theta$
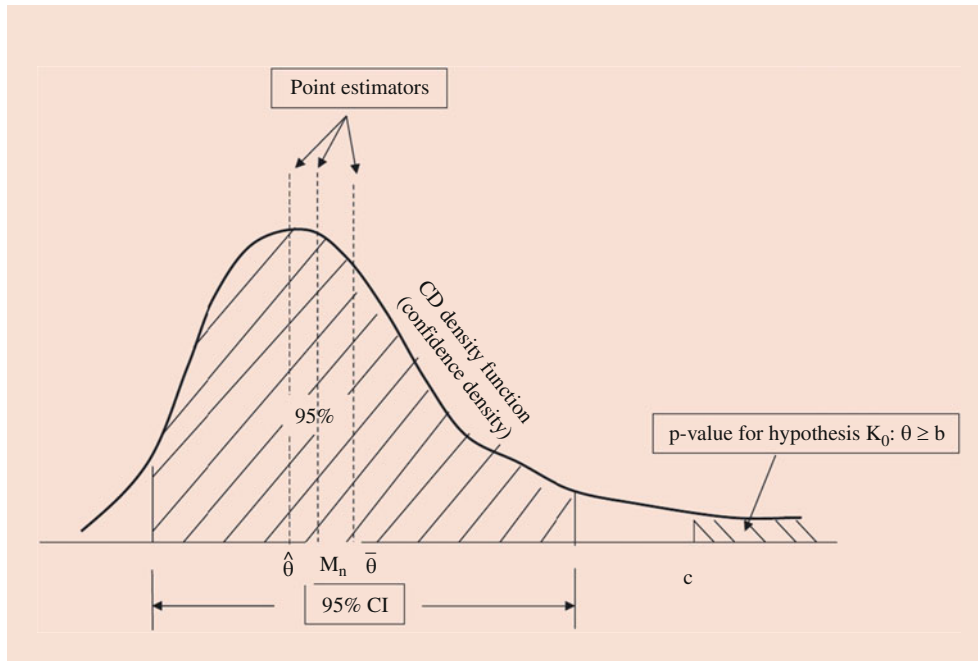
Fig. 29.3 A graphical illustration of CD-based inference [90]

is the confidence density function. Under some moderate conditions, these three point estimators are consistent [77, 90, 91].

To further understand these three types of estimators, the median $\widetilde{\theta}_n$ is an unbiased estimator with $\Pr_{\theta_0}(\widetilde{\theta}_n \leq \theta_0) = \Pr_{\theta_0}(1/2 \leq H_n(\theta_0)) = 1/2$. The mean $\bar{\theta}_n$ can be viewed as a frequentist analog of Bayesian estimator under the squared loss function. The mode $\widehat{\theta}_n$ matches with the maximum likelihood estimator if the confidence density is from a normalized likelihood function [90].

Confidence Interval As discussed in Sect. 29.2.1, in a confidence curve, a line across the y-axis of the significance level α intersects with the confidence curve at two points, and these two points correspond to an $1 - \alpha$ level, equal-tailed, two-sided confidence interval for θ, i.e., $(H_n^{-1}(\alpha/2), H_n^{-1}(1 - \alpha/2))$. Furthermore, $(-\infty, H_n^{-1}(1-\alpha)]$ and $[H_n^{-1}(\alpha), \infty)$ are one-sided $1-\alpha$ level confidence intervals for the parameter θ.

Hypothesis Testing From a CD, one can obtain p-values for various hypothesis testing problems. The natural thinking is to measure the support that $H_n(\cdot)$ lends to a null hypothesis [29]. Xie and Singh [90] summarized making inference for hypothesis testing from a CD in the following theorem.

Theorem 29.2.1 *(i) For the one-sided test* $K_0 : \theta \in C$ *versus* $K_1 : \theta \in C^c$, *where c denotes the complementary set and C is an interval of the type of* $C_l = (-\infty, b]$ *or* $C_u = [b, \infty)$, *we have* $\sup_{\theta \in C} \Pr_\theta(p(C) \leq \alpha) = \alpha$, *and* $p(C) = H_n(C)$ *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* $\Pr_{\theta=b}(2 \min\{p(C_l), p(C_u)\} \leq \alpha) = \alpha$, *and* $2 \min\{p(C_l), p(C_u)\} = 2 \min\{H_n(b), 1 - H_n(b)\}$ *is the p-value of the corresponding test.*

Example 29.2.7 ([90]) Consider Example 29.2.2 again. A CD for θ is $H_n = F_{t_{n-1}}(\frac{\sqrt{n}(\mu - \bar{x}_n)}{s_n})$. For a one-sided test $K_0 : \mu \leq b$ versus $K_1 : \mu > b$, its support on the null set $C = (-\infty, b]$ is

$$p(C) = p((-\infty, b]) = H_n(b) = F_{t_{n-1}}(\sqrt{n}(b - \bar{x}_n)/s_n).$$

This is the same p-value using the one-sided t-test. For a two-sided test $K_0 : \theta = b$ versus $K_1 : \theta \neq b$, the null set $C = \{b\}$. We would like to measure the supports of two alternative sets $p(C_l^c)$ and $p(C_u^c)$. The rejection region is defined as $\{x : 2 \max\{p(C_l^c), p(C_u^c)\} \geq 1 - \alpha\}$, i.e.,

$$\{x : 2\min\{p(C_l), p(C_u)\} \leq \alpha\} = \{x : 2\min\{H_n(b),$$
$$1 - H_n(b)\} \leq \alpha\}. \tag{29.2}$$

Under K_0 with $\theta = b$, $2\min\{p(C_l), p(C_u)\} = 2\min\{H_n(b), 1 - H_n(b)\} \sim U(0, 1)$ by the definition of a CD. Thus,

$$\Pr_{\theta=b}(2\min\{p(C_l), p(C_u)\} \leq \alpha) = \Pr_{\theta=b}(2\min\{H_n(b),$$
$$1 - H_n(b)\} \leq \alpha) = \alpha$$

and the reject region (29.2) corresponds to a level α test. Again, the p-value $2\min\{p(C_l), p(C_u)\}$ is the standard p-value from a two-sided t-test.

29.2.3 Combination of CDs for Fusion Learning

One of the important applications of CD development is on fusion learning, which synthesizes information from disparate sources with deep implications for meta-analysis [12, 15, 40, 51–53, 72, 75, 77, 81, 91, 92]. Fusion learning aims to combine inference results obtained from different data sources to achieve a more efficient overall inference result. CD-based fusion learning applies even when inference results are derived from different tests or different paradigms, i.e., Bayesian, fiducial, and frequentist (BFF).

The combination of CD can be considered as a unified framework for fusion learning. Suppose there are k independent studies that are dedicated to estimate a common parameter of interest θ. We assume that we have a CD $H^i(\cdot)$ for θ for the sample x_i of the i-th study. Singh et al. [77] proposed a general recipe for combining these k independent CDs:

$$H^c(\theta) \equiv G_c\{g_c(H^1(\theta), \ldots, H^k(\theta))\}, \qquad (29.3)$$

where g_c is a given continuous function on $[0, 1]$ which is nondecreasing in each coordinate, the function G_c is determined by the monotonic function g_c with $G_c(t) = \Pr(g_c(U_1, \ldots, U_k) \leq t)$, and $U_1, \ldots, U_k$ are independent uniform random variables. The function $H_c(\cdot)$ contains information from all k samples and is referred to as a combined CD for the parameter θ. Furthermore, the CD obtained by Eq. (29.3) does not require any information regarding how the input CDs are obtained.

A special class of the general combining framework (29.3) plays a prominent role in unifying many modern meta-analysis approaches. The choice of the function g_c for this special class is

$$g_c(u_1, \ldots, u_k) = w_1 F^{-1}(u_1) + \cdots + w_k F^{-1}(u_k), \quad (29.4)$$

where $F(\cdot)$ is a given cumulative distribution function and $w_i \geq 0$ with at least one $w_i \neq 0$ are generic weights for the combination rule. Generally, there are two types of weights: fixed weights to improve the efficiency of combination and adaptive weights based on data.

As shown in [91], it is remarkable that by choosing different g_c functions, all the classic approaches of combining p-values including Fisher, Normal (Stouffer), Min (Tippett), Max, and Sum methods [55] and all the five model-based meta-analysis estimators described in [67] including the maximum likelihood method and Bayesian approach under fixed-effects model, method of moment estimators, restricted maximum likelihood method, and Bayesian estimator with a normal prior under random-effects model, can all be obtained through a CD combination framework. Furthermore, it was shown in [94] that Mantel-Haenszel and Peto methods as well as Tian et al.'s method of combining confidence intervals [81] for meta-analysis of 2×2 tables can also all be obtained through a CD combination framework. An R-package "gmeta" developed by [95] implements the CD combining framework for fusion learning including classical p-value combination methods from [55], meta-analysis estimators with both fixed-effects and random-effects models, and many other approaches.

Fusion learning under the framework of combining CD provides an extensive and powerful tool for synthesizing information from diverse data sources. This approach has sound theoretical support and has been applied to many practical situations including robust fusion learning [91], exact fusion learning for discrete data [52, 81], fusion learning for heterogeneous studies [53], nonparametric fusion learning [15, 51], split-conquer-combine approach [12], individualized fusion learning (*i*-fusion) [75], etc. We refer to [13] for more detailed discussions.

29.2.4 Multivariate CDs

A simultaneous CD for vector parameters can sometimes be difficult to define [72], especially on how to define a multivariate CD in the exact sense in some non-Gaussian settings to ensure that their marginal distributions are CDs for the corresponding single parameter. We consider the Behrens-Fisher problem of testing for the equality of means from two multivariate normal distributions when the covariance matrices are unknown and possibly not equal. A joint CD of the two population means (μ_1, μ_2) has a joint density of the form

$$f_1\left(\frac{\mu_1 - \bar{x}_1}{s_1/\sqrt{n_1}}\right) f_2\left(\frac{\mu_2 - \bar{x}_2}{s_2/\sqrt{n_2}}\right) / \left(s_1 s_2 \sqrt{n_1 n_2}\right),$$

where f_i is the density function for the student t-distribution with $n_i - 1$ degrees of freedom, $i = 1, 2$. The marginal distribution of $\mu_1 - \mu_2$ is only an asymptotic CD but not a CD in the exact sense.

The good news in the multidimensional case is that under asymptotic settings or wherever bootstrap theory applies, one can still work with multivariate CDs [90]. When no analytic confidence curve for the parameter vector θ of interest is available, the product method of [4] can be used if confidence curves are available for each component of the vector [72]. Additionally, if we only consider center-outward confidence regions instead of all Borel sets in the $p \times 1$ parameter space, the central-CDs considered in [78] and the confidence net considered in [71] offer coherent notions of multivariate CDs in the exact sense [90].

There are many approaches to obtain CDs. One way is normalizing a likelihood function curve with respect to its

parameters so that the area underneath the curve is one. The normalized likelihood function is typically a density function. For instance, under some mild conditions, Fraser and McDunnough [32] show that this normalized likelihood function is the normal density function of an asymptotic CD. Other ways like bootstrap distributions and p-value functions also often provide valid CDs. Finally, CDs and fiducial distributions have been always linked since their inception. The class of fiducial inference provides another systematic way to obtain CDs and we will further discuss fiducial inference in the next section.

29.3 Fiducial Inference

CD can be somehow viewed as "the Neymanian interpretation of Fisher's fiducial distributions" [74]. From the definition of CD and fiducial distribution, we may consider the fiducial distribution as one special type of CD, though the CD looks at the problem of obtaining an inferentially meaningful distribution on the parameter space from a pure frequentist point of view [90]. Nevertheless, fiducial inference provides a systematic way to obtain a CD, and its development provides a rich class of literature for CD inference. We briefly review fiducial inference and its recent developments in this section.

29.3.1 Fiducial Inference

R.A. Fisher introduced the idea of fiducial probability and fiducial inference [28] as a potential replacement of the Bayesian posterior distribution. Although he discussed fiducial inference in several subsequent papers, there appears to be no rigorous definition of a fiducial distribution for a vector parameter. The basic idea of the fiducial argument is switching the role of data and parameters to introduce the distribution on the parameter space. This obtained distribution then summarizes our knowledge about the unknown parameter. Since the mid-2000s, there has been a renewed interest in modern modifications of fiducial inference. The common approaches for these modifications rely on a definition of inferentially meaningful probability statements about subsets of the parameter space without introducing any prior information.

These modern approaches include generalized fiducial inference [37, 41], Dempster-Shafer theory [22, 24], and inferential models [56, 61]. Objective Bayesian inference, which aims at finding nonsubjective model-based priors, can also be seen as addressing the same question. Examples of recent breakthroughs related to reference prior and model selection are [2, 7, 8]. Another related approach is based on higher-order likelihood expansions and implied data-dependent

priors [30, 31, 33–36]. There are many more references that interested readers can find in [41].

29.3.2 Generalized Fiducial Distribution

Generalized fiducial inference, motivated by [83, 84], has been at the forefront of the modern fiducial revival. Generalized fiducial inference defines a data-dependent measure on the parameter space by using an inverse of a deterministic data generating equation without the use of Bayes theorem.

Motivated by Fisher's fiducial argument, generalized fiducial inference begins with expressing the relationship between the data Y and the parameters θ as

$$Y = G(U, \theta), \tag{29.5}$$

where $G(\cdot, \cdot)$ is a deterministic function termed as the data generating equation and U is the random component of this data generating equation whose distribution is independent of parameters and completely known.

The data Y are created by generating a random variable U and plugging it into the data generating equation (29.5). For example, a single observation from $N(\mu, 1)$ distribution can be written as $Y = \mu + U$, where $\theta = \mu$ and U is $N(0, 1)$ random variable.

Fisher's original fiducial argument only addresses the simple case where the data generating equation (29.5) can be inverted and the inverse $Q_y(u) = \theta$ exists for any observed y and for any arbitrary u. One can define the fiducial distribution for θ as the distribution of $Q_y(U^\star)$ where $U^\star$ is an independent copy of U. Equivalently, a sample from the fiducial distribution of θ can be obtained by first generating $U_i^\star$, and then let $\theta_i^\star = Q_y(U_i^\star)$, $i = 1, \ldots, n$. Point estimation and confidence intervals for θ can be obtained based on this sample. In the $N(\mu, 1)$ example, $Q_y(u) = y - u$ and the fiducial distribution is therefore the distribution of $y - U^\star \sim N(y, 1)$.

In the case of no θ satisfying Eq. (29.5), Hannig [37] proposed to use the distribution of U conditional on the event $\{u : y = G(u, \theta), \text{ for some } \theta\}$. Hannig et al. [41] generalized this approach and proposed an attractive definition of generalized fiducial distribution (GFD) through a weak limit.

Definition 29.3.1 A probability measure on the parameter space Θ is called a GFD if it can be obtained as a weak limit

$$\lim_{\epsilon \to 0} \left[\arg\min_{\theta^\star} \|y - G(U^\star, \theta^\star)\| \;\middle|\; \min_{\theta^\star} \|y - G(U^\star, \theta^\star)\| \leq \epsilon \right]. \tag{29.6}$$

Hannig et al. [41] pointed out a close relationship between GFD and approximate Bayesian computations (ABC) [3].

In an idealized ABC, one first generates an observation θ^* from the prior, then generates a new sample using a data generating equation $y^\star = G(U^\star, \theta^\star)$, and compares the generated data with the observed data y. If the observed and generated datasets are close, i.e., $\|y - y^\star\| \leq \epsilon$, the generated $\theta^\star$ is accepted; otherwise it is rejected and the procedure is repeated. On the other hand, as for GFD, one first generates $U^\star$, finds a best fitting $\theta^\star = \arg\min_{\theta^\star} \|y - G(U^\star, \theta^\star)\|$, computes $y^\star = G(U^\star, \theta^\star)$, again accepts $\theta^\star$ if $\|y - y^\star\| \leq \epsilon$, and rejects otherwise. In either approach an artificial dataset $y^\star = G(U^\star, \theta^\star)$ is generated and compared to the observed data. The main difference is that the Bayes posterior simulates the parameter $\theta^\star$ from the prior, while GFD uses the best-fitting parameter.

Fiducial distributions often have good frequentist properties, and corresponding fiducial confidence intervals often give asymptotically correct coverage [37, 41]. In addition, fiducial distribution is a data-dependent measure on the parameter space and thereby a CD. Xie and Singh [90] described the relation between the concepts of CD and fiducial distributions using an analogy in point estimation: A CD is analogous to a consistent estimator and a fiducial distribution is analogous to a maximum likelihood estimator. In the context of point estimation, a consistent estimator does not have to be a maximum likelihood estimator. But under some regularity conditions, the maximum likelihood estimator typically provides a standard procedure to obtain a consistent estimator. In the context of distribution estimator, a CD does not have to be a fiducial distribution. However, under suitable conditions, a fiducial distribution often has good frequentist properties and thus a CD.

29.3.3 A User-Friendly Formula for GFD

While Definition (29.6) for GFD is conceptually and mathematically appealing, it is not clear how to compute the limit in most of practical situations. The following theorem proposed by [41] provides a computational tool.

Theorem 29.3.1 *Under certain assumptions, the limiting distribution in (29.6) has a density*

$$r(\theta|y) = \frac{f(y, \theta)J(y, \theta)}{\int_\Theta f(y, \theta')J(y, \theta')\, d\theta'}, \qquad (29.7)$$

where $f(y, \theta)$ is the likelihood and the function

$$J(y, \theta) = D\left(\left.\frac{d}{d\theta}G(u, \theta)\right|_{u=G^{-1}(y,\theta)}\right). \qquad (29.8)$$

If (i) $n = p$, then $D(A) = |\det A|$. Otherwise the function $D(A)$ depends on the norm used; (ii) the l_∞

norm gives $D(A) = \sum_{\mathbf{i}=(i_1,\ldots,i_p)} |\det(A)_{\mathbf{i}}|;$[1] *(iii) under an additional assumption stated in [41], the l_2 norm gives* $D(A) = (\det A^\top A)^{1/2}.$

Hannig et al. [41] recommended using (ii) for practitioners. A nice property of GFD is that GFD is invariant under smooth re-parameterizations. This property follows directly from (29.6), since for an appropriate selection of minimizers and any one-to-one function $\theta = \phi(\eta)$,

$$\phi\left(\arg\min_{\eta^\star} \|y - G(U^\star, \phi(\eta^\star))\|\right) = \arg\min_{\theta^\star} \|y - G(U^\star, \theta^\star)\|.$$

Note that GFD could change with transformations of the data generating equation. Assume that the observed dataset has been transformed by a one-to-one smooth transformation $Z = T(Y)$. By the chain rule, the GFD based on this new data generating equation and observed data $z = T(y)$ is the density (29.7) with the Jacobian function

$$J_T(z, \theta) = D\left(\left.\frac{d}{dy}T(y) \cdot \frac{d}{d\theta}G(u, \theta)\right|_{u=G^{-1}(y,\theta)}\right), \qquad (29.9)$$

where for simplicity we write y instead of $T^{-1}(z)$.

29.3.4 Examples of GFD

In this section we will consider two examples, linear regression and uniform distribution. In the first case, the GFD is the same as Bayes posterior with respect to the independence Jeffreys prior, while in the second case, the GFD is not a Bayes posterior with respect to any prior (that is not data dependent).

Linear Regression [41] We consider a generalized fiducial approach to regression problem. We express linear regression via the data generating equation,

$$Y = G(U, \theta) = X\beta + \sigma U,$$

where Y is the dependent variables, X is the design matrix, $\theta = (\beta, \sigma)$ are the unknown parameters, and U is a random vector with known density $f(u)$ independent of θ and X. Note that $\frac{d}{d\theta}G(U, \theta) = (X, U)$ and $U = (y - X\beta)/\sigma$; the Jacobian in (29.9) using the l_∞ norm simplifies to

---

[1] In (ii) the sum spans over $\binom{n}{p}$ of p-tuples of indexes $\mathbf{i} = (1 \leq i_1 < \cdots < i_p \leq n)$. For any $n \times p$ matrix A, the sub-matrix $(A)_{\mathbf{i}}$ is the $p \times p$ matrix containing the rows $\mathbf{i} = (i_1, \ldots, i_p)$ of A.

$$J_\infty(y, \theta) = \sigma^{-1} \sum_{\substack{i=(i_1,\dots,i_p) \\ 1 \le i_1 < \cdots < i_p \le n}} |\det (X, Y)_i|,$$

and the density of GFD is

$$r(\beta, \sigma|y) \propto \sigma^{-n-1} f((Y - X\beta)/\sigma).$$

The fiducial solution is the same as the Bayesian solution using Jeffreys prior [5]. Furthermore, by a simple calculation, the Jacobian with l_2 norm differs from $J_\infty(y, \theta)$ only by a constant; the GFD remains unchanged.

GFD in Irregular Models [41] We consider an irregular model $U(a(\theta) - b(\theta), a(\theta) + b(\theta))$. The reference prior for

$$r_1(\theta|y) \propto \frac{a'(\theta) - a(\theta)\{\log b(\theta)\}' + \bar{y}_n\{\log b(\theta)\}'}{b(\theta)^n} I_{\{a(\theta)-b(\theta)<y_{(1)} \ \& \ a(\theta)+b(\theta)>y_{(n)}\}}.$$

Consider an alternative fiducial solution, which constructs the GFD based on the minimal sufficient and ancillary statistics $Z = \{h_1(Y_{(1)}), h_2(Y_{(n)}), (Y - Y_{(1)})/(Y_{(n)} - Y_{(1)})\}^\top$, where

$$J_2(y, \theta) = (w_1 + w_2)\left[a'(\theta) - a(\theta)\{\log b(\theta)\}' + \frac{w_1 y_{(1)} + w_2 y_{(n)}}{w_1 + w_2}\{\log b(\theta)\}'\right],$$

$$r_2(\theta|y) \propto \frac{I_{\{a(\theta)-b(\theta)<y_{(1)} \ \& \ a(\theta)+b(\theta)>y_{(n)}\}}}{[(w_1 + w_2)[a'(\theta) - a(\theta)\{\log b(\theta)\}'] + (w_1 y_{(1)} + w_2 y_{(n)})\{\log b(\theta)\}']^{-1} b(\theta)^n},$$

where $w_1 = h_1'(y_{(1)})$ and $w_2 = h_2'(y_{(n)})$.

Hannig et al. [41] performed extensive simulation studies for a particular case $U(\theta, \theta^2)$ comparing GFD to the Bayesian posteriors with the reference prior $\pi(\theta) = \frac{(2\theta-1)}{\theta(\theta-1)} e^{\psi\left(\frac{2\theta}{2\theta-1}\right)}$ [7][2] and flat prior $\pi(\theta) = 1$. The simple GFD, the alternative GFD, and the reference prior Bayes posterior maintain nominal coverage for all parameter settings. However, the flat prior Bayes posterior does not have a satisfactory coverage, with the worst departures from nominal coverage for small sample size and large parameter θ.

Nonparametric Fiducial Inference with Right-Censored Data [17] Let failure times X_i $(i = 1, \dots, n)$ follow the true distribution function F_0 and censoring times C_i $(i = 1, \dots, n)$ have the distribution function R_0. We treat the situation when failure and censoring times are independent and unknown.

this model has been shown complex in Theorem 8 from [7]. Considering GFD approach, we first express the observed data by the following data generating equation:

$$Y_i = a(\theta) + b(\theta)U_i, \quad U_i \overset{i.i.d.}{\sim} U(-1, 1).$$

By simple algebra,

$$\frac{d}{d\theta}G(u, \theta) = a'(\theta) + b'(\theta)U \text{ with } U = b^{-1}(\theta)(Y - a(\theta)).$$

If $a'(\theta) > |b'(\theta)|$, (29.8) simplifies to

$$J_1(y, \theta) = n[a'(\theta) - a(\theta)\{\log b(\theta)\}' + \bar{y}_n\{\log b(\theta)\}'],$$

and the GFD is

$Y_{(1)}, Y_{(n)}$ are order statistics, $h_1^{-1}(\theta) = EY_{(1)} = a(\theta) - b(\theta)(n-1)/(n+1)$ and $h_2^{-1}(\theta) = EY_{(n)} = a(\theta) + b(\theta)(n-1)/(n+1)$. By a simple calculation,

Suppose we observe right-censored data $\{y_i, \delta_i\}$ $(i = 1, \dots n)$, where $y_i = x_i \wedge c_i$ is the minimum of x_i and c_i, $\delta_i = I\{x_i \le c_i\}$ denotes censoring indicator.

Consider the following data generating equation:

$$Y_i = F^{-1}(U_i) \wedge R^{-1}(V_i), \quad \Delta_i = I\{F^{-1}(U_i) \le R^{-1}(V_i)\}$$

$$(i = 1, \dots n),$$

where U_i, V_i are independent and identically distributed $U(0, 1)$.

For a failure event $\delta_i = 1$, we have full information about failure time x_i, i.e., $x_i = y_i$, and partial information about censoring time c_i, i.e., $c_i \ge y_i$. Thus,

$$F^{-1}(u_i) = y_i \Longleftrightarrow F(y_i) \ge u_i, F(y_i - \epsilon) < u_i \text{ for any } \epsilon > 0.$$

For a censored event $\delta_i = 0$, we only know partial information about x_i, i.e., $x_i > y_i$, and full information on c_i, i.e., $c_i = y_i$. Similarly,

---

[2] $\psi(x)$ is the digamma function defined by $\psi(z) = \frac{d}{dz}\log(\Gamma(z))$ for $z > 0$, where Γ is gamma function.

$$F^{-1}(u_i) > y_i \iff F(y_i) < u_i,$$

$$R^{-1}(v_i) = y_i \iff R(y_i) \geq v_i, R(y_i - \epsilon) < v_i \text{ for any } \epsilon > 0.$$

The complete inverse map of the data generating equation is

$$Q^F(y, \delta, u) = \left\{ F : \begin{cases} F(y_i) \geq u_i, F(y_i - \epsilon) < u_i \text{ for any } \epsilon > 0 & \text{for all } i \text{ such that } \delta_i = 1 \\ F(y_j) < u_j & \text{for all } j \text{ such that } \delta_j = 0 \end{cases} \right\}, \quad (29.11)$$

and $Q^R(y, \delta, v)$ is analogous.

Let (U^*, V^*) be an independent copy of (U, V). Because the inverse (29.10) separates into a Cartesian product, and of the fact that U^* and V^* are independent, the marginal fiducial distribution for the failure distribution function F is

$$Q^F(y, \delta, U^*) \mid \{Q^F(y, \delta, U^*) \neq \emptyset\}.$$

Figure 29.4 from [17] demonstrates the survival function representation of $Q^F(y, \delta, u)$, as defined in Eq. (29.11), for one dataset with $n = 8$ observations of X following *Weibull*(20, 10) censored by Z following *Exp*(20). Each of the panels corresponds to a different value of u, where each u is a realization of U^*. Any survival function lying between the upper red and the lower black fiducial survival functions corresponds to an element of the closure of $Q^F(y, \delta, u)$. The technical details of sampling refer to Algorithm 1 in [17]. The corresponding fiducial-based confidence intervals proposed in [17] maintain coverage in situations where asymptotic methods often have substantial coverage problems. Furthermore, as also shown in [17], the average length of their log-interpolation fiducial confidence intervals is often shorter than the length of confidence intervals for competing methods that maintain coverage. As pointed by [80], it would also be interesting to consider other choices of fiducial samples such as monotonic spline interpolation.

GFDs for Discrete Distributions [41] Let Y be a random variable with distribution function $F(y|\theta)$. Assume there is $\mathcal{Y}$ so that $P_\theta(Y \in \mathcal{Y}) = 1$ for all θ, and for each fixed $y \in \mathcal{Y}$, the distribution function is either a nonincreasing function of θ, spanning the whole interval $(0, 1)$, or a constant equal to 1; the left limit $F(y_-|\theta)$ is also either a nonincreasing function of θ spanning the whole interval $(0, 1)$ or a constant equal to 0.

Define $F^-(a|\theta) = \inf\{y : F(y|\theta) \geq a\}$. It is well known [11] that if $U \sim U(0,1)$, $Y = F^-(U|\theta)$ has the correct distribution and we use this association as a data generating equation. It follows that both $Q_y^+(u) = \sup\{\theta : F(y|\theta) = u\}$ and $Q_y^-(u) = \inf\{\theta : F(y_-|\theta) = u\}$ exist and satisfy $F(y|Q_y^+(u)) = u$ and $F(y_-|Q_y^-(u)) = u$. Consequently,

$$Q^{F,R}(y, \delta, u, v) = \bigcap_i Q_{\delta_i}^{F,R}(y_i, u_i, v_i) = Q^F(y, \delta, u) \times Q^R(y, \delta, v), \quad (29.10)$$

where

$$P(Q_y^+(u) \leq t) = 1 - F(y|t) \text{ and}$$

$$P(Q_y^-(u) \leq t) = 1 - F(y_-|t).$$

Note that for all $u \in (0, 1)$, the function $F^-(u|\theta)$ is nonincreasing in θ and the closure of the inverse image $\bar{Q}_y(u) = \{Q_y^-(u), Q_y^+(u)\}$. The half-corrected GFD has distribution function

$$R(\theta|y) = 1 - \frac{F(y|\theta) + F(y_-|\theta)}{2}.$$

If either of the distribution functions is constant, we interpret it as a point mass at the appropriate boundary of the parameter space. Analogous argument shows that if the distribution function and its left limit were nondecreasing in θ, the half-corrected GFD would have distribution function

$$R(\theta|y) = \frac{F(y|\theta) + F(y_-|\theta)}{2}.$$

Hannig et al. [41] provide a list of the half-corrected GFDs for three well-known discrete distributions. Let Beta(0, $n +$ 1) and Beta($x + 1$, 0) denote the degenerate distributions on 0 and 1, respectively. Let $\Gamma(0, 1)$ denote the degenerate distribution on 0:

- $X \sim$ Binomial(m, p) with m known. GFD is the mixture of Beta($x + 1, m - x$) and Beta($x, m - x + 1$) distributions [37].
- $X \sim$ Poisson(λ). GFD is the mixture of $\Gamma(x + 1, 1)$ and $\Gamma(x, 1)$ distributions [22].
- $X \sim$ Negative Binomial(r, p) with r known. GFD is the mixture of Beta($r, x - r + 1$) and Beta($r, x - r$) distributions [38].

Model Selection via GFD [41] Hannig and Lee [39] introduced model selection into the generalized fiducial inference paradigm in the context of wavelet regression. Two important ingredients are needed for fiducial model selection: (1) include the choice of model as one of the parameters; (2) include penalization in the data generating equation.

Consider a finite collection of models $\mathcal{M}$. The data generating equation is
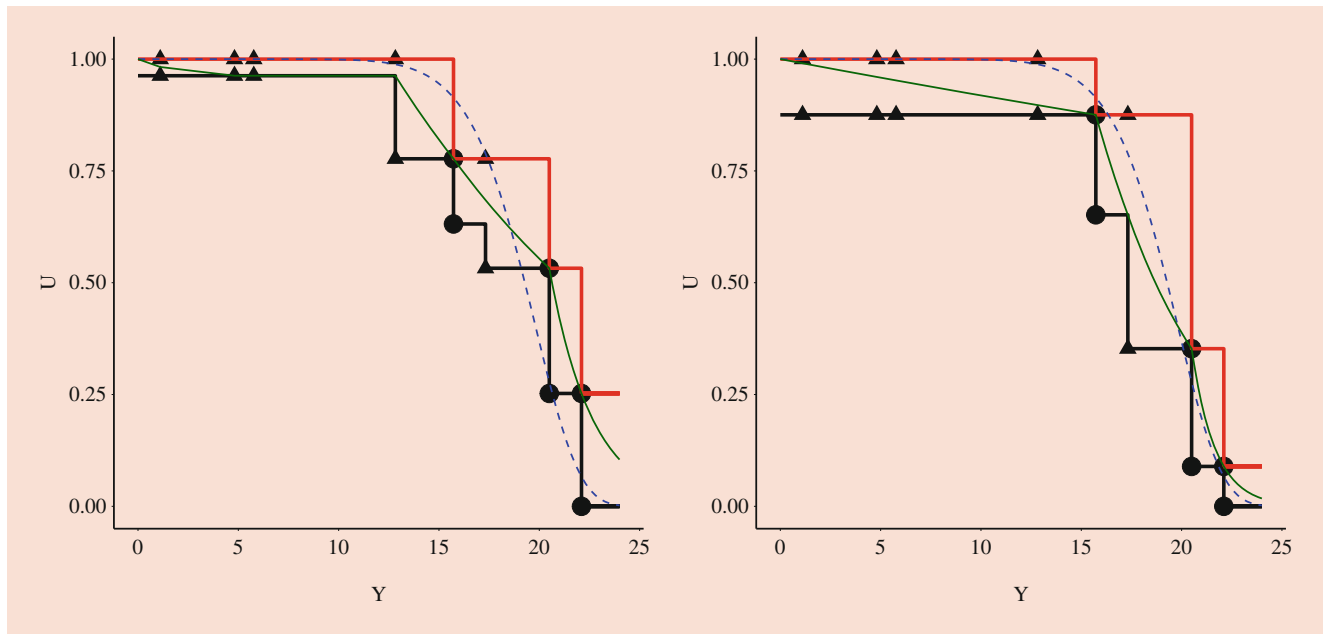
Fig. 29.4 Two realizations of fiducial curves for a sample of size 8 from *Weibull*(20, 10) censored by *Exp*(20) [17]. Here fiducial curves refer to Monte Carlo samples S_i^L, S_i^U, and S_i^I ($i = 1, 2$) from the GFD. The red and black curves are corresponding realizations of the upper and lower fiducial survival functions. The green curve is the log-linear interpolation type of survival functions. The circle points denote failure observations. The triangle points denote censored observations. The dashed blue curve is the true survival function of *Weibull*(20, 10)

$$Y = G(M, \theta_M, U), \qquad M \in \mathcal{M}, \ \theta_M \in \Theta_M, \qquad (29.12)$$

where Y is the observation, M is the model considered, θ_M includes the parameters associated with model M, and U is a random vector of with fully known distribution independent of any parameters. Hannig and Lee [39] proposed a novel way of adding a penalty into the fiducial model selection. In particular, for each model M, they proposed to augment the data generating equation (29.12) by

$$0 = P_k, \quad k = 1, \ldots, \min(|M|, n), \qquad (29.13)$$

where P_k are independent and identically distributed continuous random variables independent of U with $f_P(0) = q$ and q is a constant determined by the penalty. Hannig and Lee [39] recommended using $q = n^{-1/2}$ as the default penalty. Note that the number of additional equations is the same as the number of unknown parameters in the model. As we never actually observe the outcomes of the extra data generating equations, we will select their values as $p_i = 0$.

For the augmented data generating equation, we have the following theorem from [41]. The quantity $r(M|y)$ can be used for inference in the usual way. For example, fiducial factor, the ratio $r(M_1|y)/r(M_2|y)$, can be used in the same way as a Bayes factor, as discussed in [6] in the context of Bayesian model selection.

Theorem 29.3.2 ([41]) *Suppose $|M| \leq n$ and certain assumptions hold; the marginal generalized fiducial probability of model M is*

$$r(M|y) = \frac{q^{|M|} \int_{\Theta_M} f_M(y, \theta_M) J_M(y, \theta_M) \, d\theta_M}{\sum_{M' \in \mathcal{M}} q^{|M'|} \int_{\Theta_{M'}} f_{M'}(y, \theta_{M'}) J_{M'}(y, \theta_{M'}) \, d\theta_{M'}}, \qquad (29.14)$$

where $f_M(y, \theta_M)$ is the likelihood and $J_M(y, \theta_M)$ is the Jacobian function computed using (29.9) for each fixed model M.

For more details on the use of fiducial model selection, see [39] and [43].

29.4 Applications and Numerical Examples

29.4.1 CD-Based Inference

Two-Parameter Exponential Distribution Inference procedures based on the two-parameter exponential model, $Exp(\mu, \sigma)$, are extensively used in several areas of statistical practice, including survival and reliability analysis. The probability distribution function and cumulative distribution function of a random variable $X \sim Exp(\mu, \sigma)$ are given, respectively, by

$$f(x) = \frac{1}{\sigma} \exp \left\{ -\frac{x - \mu}{\sigma} \right\},$$

$$F(x) = \begin{cases} 1 - \exp \left\{ -\frac{x-\mu}{\sigma} \right\} & \text{if } x > \mu, \\ 0 & \text{if } x \leq \mu, \end{cases}$$

and survival function (also known as reliability function) is $S(x) = 1 - F(x)$. The inference problem of interest is to obtain confidence intervals (sets) of μ, σ and $S(t)$ at a given $t > 0$.

Let $X_{(1)}, \ldots, X_{(k)}$ be the k ($k > 1$) smallest observations among $X_1, \ldots, X_n$. Then the maximum likelihood estimator of μ and σ are

$$\widehat{\mu} = X_{(1)}, \quad \text{and} \quad \widehat{\sigma} = \frac{1}{k}\left\{ \sum_{i=1}^{k} X_{(i)} + (n-k)X_{(k)} - nX_{(1)} \right\}.$$

It turns out that $\widehat{\mu}$ and $\widehat{\sigma}$ are independent and they follow the distributions

$$U = 2n(\widehat{\mu} - \mu)/\sigma \sim \chi^2(2), \quad V = 2k\widehat{\sigma}/\sigma \sim \chi^2(2k-2),$$
(29.15)

respectively. Here $\chi^2(m)$ is the chi-square distribution with degree of freedom m. We provide below a simple CD-based method to answer the inference problem of interest.

From Eq. (29.15), we have

$$\frac{n(\widehat{\mu} - \mu)}{k\widehat{\sigma}} = \frac{U/2}{V/(2k-2)} \sim F(2, 2k-2),$$

where $F(a, b)$ is the F-distribution with degrees of freedom a and b. By the pivot-based CD construction method [78, p134], a CD for μ is $H_1(\mu) = 1 - F_{F(2,2k-2)}(\frac{n(\widehat{\mu}-\mu)}{k\widehat{\sigma}})$, where $F_{F(2,2k-2)}$ is the cumulative distribution function of $F(2, 2k-2)$-distribution. Similarly, a CD for σ is $H_2(\sigma) = 1 - F_{\chi^2(2k-2)}(\frac{2k\widehat{\sigma}}{\sigma})$, where $F_{\chi^2(2k-2)}$ is the cumulative distribution function of $\chi^2(2k-2)$-distribution. Inferential statements regarding μ and σ, including confidence intervals and testing results, can be obtained from these two CDs. Coverage rates and test errors obtained from these two CDs are exact.

We can also consider the inference for (μ, σ) jointly. Here, we introduce a simulation-based approach. Let $U^* \sim \chi^2(2)$ and $V^* \sim \chi^2(2k-2)$ be two independently simulated random numbers. Define

$$\xi^* = \widehat{\mu} - \frac{k\widehat{\sigma}}{n}\frac{U^*}{V^*} \quad \text{and} \quad \zeta^* = \frac{2k\widehat{\sigma}}{V^*}.$$

Then, $\xi^*|(\widehat{\mu}, \widehat{\sigma}) \sim H_1(\mu)$ and $\zeta^*|(\widehat{\mu}, \widehat{\sigma}) \sim H_2(\sigma)$, and they are called *CD random variables* [90]. Furthermore, the underlying joint distribution of (ξ^*, ζ^*), given $(\widehat{\mu}, \widehat{\sigma})$, is a joint CD function $H_3(\mu, \sigma)$ of (μ, σ). If we simulate a large number of, say M, copies of (U^*, V^*), then we can get M copies of (ξ^*, ζ^*). In order to make inference statements about (μ, σ), we can treat these M copies of $(\xi_1^*, \zeta_1^*), \ldots, (\xi_M^*, \zeta_M^*)$ as if they were M copies of bootstrap estimators in bootstrap inference or as if they were M copies of random samples from the posterior distribution of (μ, σ) in a Bayesian inference.

Additionally, we can also use the M copies of CD random variables $(\xi_1^*, \zeta_1^*), \ldots, (\xi_M^*, \zeta_M^*)$ to obtain a pointwise confidence band for $S(t)$, $t > 0$. For each given $t > 0$, we compute $\kappa_j^*(t) = \exp\{-(t - \xi_j^*)/\zeta_j^*\}$, for $j = 1, \ldots, M$. Then $[\kappa_{[\alpha M]}^*(t), +\infty)$ and $[\kappa_{[\frac{\alpha}{2}M]}^*(t), \kappa_{[\frac{(1-\alpha)}{2}M]}^*(t)]$ are the one-sided and two-sided level-α confidence intervals of $S(t)$, respectively, where $\kappa_{[qM]}^*(t)$ is the q-th quantile of $\kappa_1^*(t), \ldots, \kappa_M^*(t)$. Now by varying t, $[\kappa_{[\alpha M]}^*(t), +\infty)$ forms a level-α lower confidence band, and $[\kappa_{[\frac{\alpha}{2}M]}^*(t), \kappa_{[\frac{(1-\alpha)}{2}M]}^*(t)]$ forms a level-α confidence band for the survival function $S(t)$.

We can show that this set of exact confidence bands derived from the CD method matches with those obtained in [69] using Tsui and Weerahandi's generalized inference approach [83], but the CD approach is very simple and more direct. Roy and Mathew [69] illustrated the 95% lower limit $\widetilde{S}(t)$ for time ranging from 150 to 2000 in Figure 1 of [69] using a real data example with 19 observations taken from [45]. The data deal with mileages for military personnel carriers that failed in service. Figure 29.5 is a similar plot for the confidence band, using our CD approach with $M = 1000$.

Data [45]:
162, 200, 271, 320, 393, 508, 539, 629, 706, 777, 884, 1008, 1101, 1182, 1463, 1603, 1984, 2355, 2880

Bivariate Normal Correlation Suppose we have the following bivariate normal distribution:

$$N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right),$$
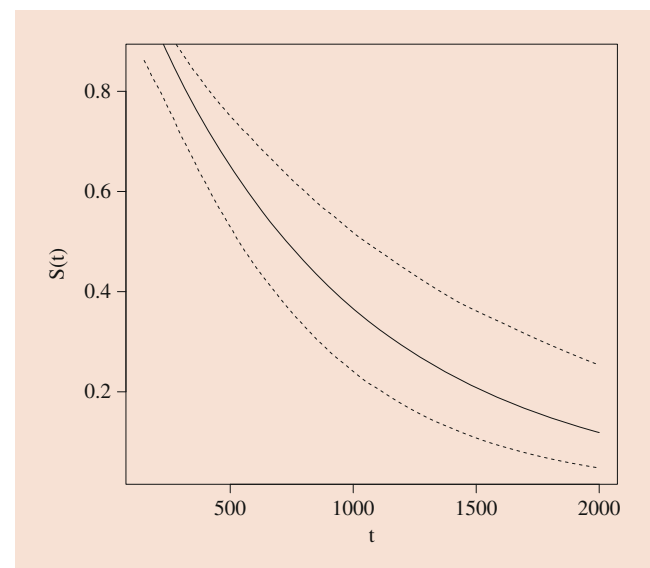

Fig. 29.5 Point estimate (solid line) and 95% confidence band (dashed line) of CD-based inference

and let ρ denote the correlation coefficient. One could use the asymptotic pivot, Fisher's Z [27, 78],

$$\frac{1}{2} \log \frac{1+r}{1-r} - \frac{1}{2} \log \frac{1+\rho}{1-\rho},$$

where r is the sample correlation. The limiting distribution of the above pivot is $N(0, \frac{1}{n-3})$. Therefore, the asymptotic CD is

$$H_n(\rho) = 1 - \Phi\left(\sqrt{n-3}\left[\frac{1}{2}\log\frac{1+r}{1-r} - \frac{1}{2}\log\frac{1+\rho}{1-\rho}\right]\right), \quad -1 \le \theta \le 1.$$

Figure 29.6 presents the CD of correlation coefficient ρ for a simulated dataset with $n = 50, \mu_1 = \mu_2 = 1, \sigma_1 = \sigma_2 = 1, \rho = 0.5$.

In addition to the above two examples, there also are recent developments of CDs on causal inference; see more applications in [54].
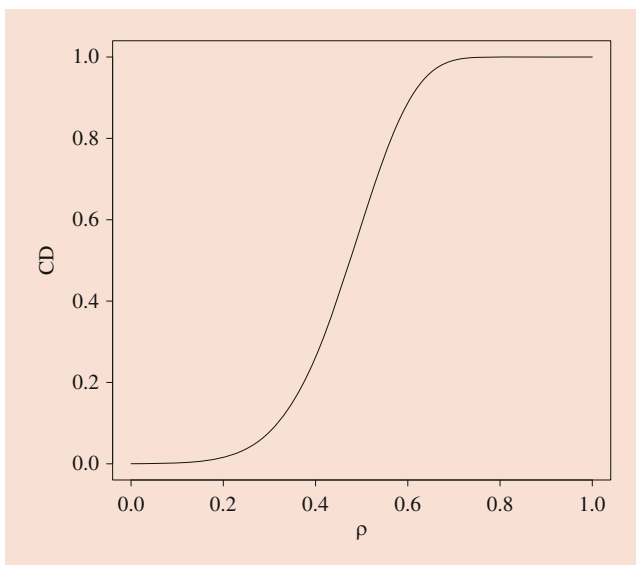
29.4.2 Nonparametric GFD-Based Inference

[17] proposed a fiducial approach to testing reliability function with an infinite dimensional parameter. Their approach does not assume a parametric distribution and is robust to model mis-specification. In [17], they considered a clinical trial of chemotherapy against chemotherapy combined with radiotherapy in the treatment of locally unresectable gastric cancer conducted by the Gastrointestinal Tumor Study Group [70]. In this trial, 45 patients were randomized to each of the 2 groups and followed for several years. The censoring percentage is 13.3% for the combined therapy group, and 4.4% for the chemotherapy group. We are interested in testing whether the two treatment groups have the same survival functions.

The Kaplan-Meier curves for these two datasets are presented in Fig. 29.7a. We notice that the two hazards appear to be crossing, which could pose a problem for some log-rank
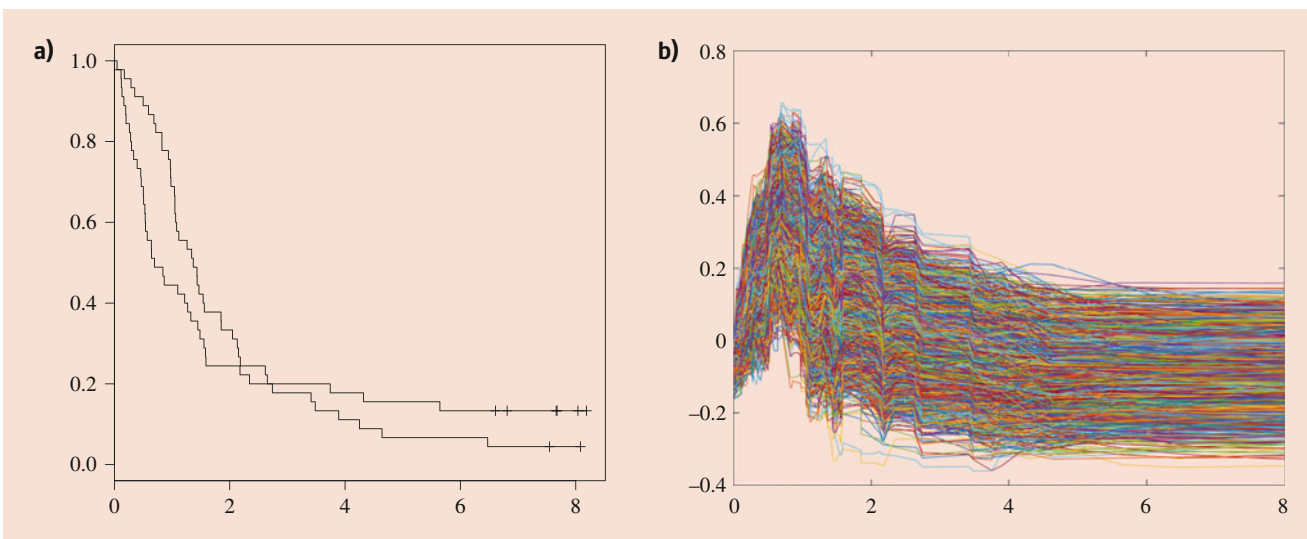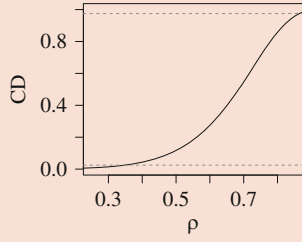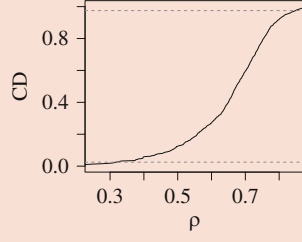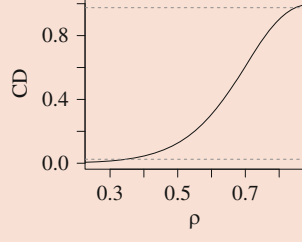

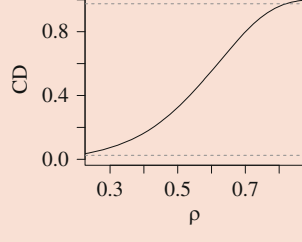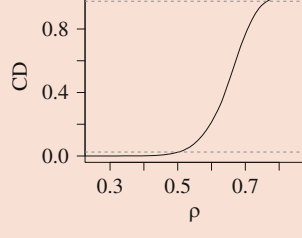**Fig. 29.6** CD of the correlation coefficient ρ

Fig. 29.7 (**a**) Kaplan-Meier estimators for two treatment groups [17]. (**b**) Difference of two sample fiducial distributions

tests. In this instance, the fiducial approach gives a small p-value 0.002. The p-values of other types of log-rank tests are reported in [17]. To explain why their proposed fiducial approach works good, they plot the sample of the difference of two fiducial distributions in Fig. 29.7b. If these two datasets are from the same distribution, 0 should be well within the sample curves. However, from Fig. 29.7b, we could see that the majority of curves are very far away from 0 on the interval [0.5, 1]. This gives strong evidence that the group

with combined therapy has significantly worse early survival outcomes.

In [17], they choose to use the sup-norm in the definition of the curvewise confidence intervals and tests. It could be possible to make the procedure more powerful by using a different (possibly weighted) norm [64]. Similarly, it might also be possible to use the choice of norm motivated by inferential models [18, 58, 61]. Besides the above example, there also are recent developments of non-

Table 29.1 Inference on correlation coefficient: combining independent bivariate normal studies

Methods	95% CI	CDs
Fisher's Z method	(0.348,0.845)	
Bootstrap BC$_a$	(0.317,0.818)	
Profile likelihood	(0.346,0.827)	
Bayes (uniform prior)	(0.188,0.790)	
Combination	(0.505,0.760)	

parametric fiducial inference on interval-censored data and Efron's empirical Bayes deconvolution; see [19, 20] for more applications.

Data [70]: (* indicates a censored event)
Combination group: 0.05 0.12 0.12 0.13 0.16 0.20 0.20 0.26 0.28 0.30 0.33 0.39 0.46 0.47 0.50 0.51 0.53 0.53 0.54 0.57 0.64 0.64 0.70 0.84 0.86 1.10 1.22 1.27 1.33 1.45 1.48 1.55 1.58 1.59 2.18 2.34 3.74 4.32 5.64 6.61* 6.81* 7.66* 7.68* 8.04* 8.19*
Chemotherapy group: 0.00 0.17 0.29 0.35 0.50 0.59 0.68 0.72 0.82 0.82 0.94 0.97 0.98 0.98 1.04 1.05 1.05 1.06 1.08 1.12 1.26 1.34 1.37 1.43 1.44 1.47 1.54 1.56 1.85 1.85 2.05 2.13 2.15 2.18 2.62 2.65 2.74 3.41 3.48 3.89 4.25 4.64 6.47 7.55* 8.08*

29.4.3 Combining Information from Multiple CDs

We use simple cluster of differentiation 4 (cd-4) count data considered in [23] to demonstrate combining information from CDs. Twenty HIV-positive subjects received an experimental antiviral drug. The cd-4 counts in hundreds were recorded for each subject at baseline and after 1 year of treatment.

We obtained the summary statistics and simulated four independent datasets from the following bivariate normal distribution:

$$N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right),$$

where $\mu_1 = 3.288$, $\mu_2 = 4.093$, $\sigma_1^2 = 0.657$, $\sigma_2^2 = 1.346$, and $\rho = 0.723$.

Suppose each study makes its own inference conclusion individually. Each dataset was analyzed by Fisher's Z method [27, 76], the bias-corrected and accelerated (BC$_a$) bootstrap [10, 21, 23], the profile likelihood approach [48], and Bayesian with uniform prior [1], respectively. One natural question we would like to ask is if we can combine the inferences from four independent studies, given that ρ is the same in all studies. The answer is yes. As introduced in Section 29.2.3, combination of CDs is a powerful inferential tool. We fused studies by combining p-values (Stouffer) [55, 95].

The results of the analysis are summarized in Table 29.1. As we can see from the table, four methods in different studies provide more or less similar results, and the combined interval is much shorter than any of the four individual intervals. In order to study the performance of the combination of CDs in this situation, we present a simulation study with 200 replications. Table 29.2 shows the

Table 29.2 Combination of four independent bivariate normal studies via CDs

Methods	Coverage	Mean length (sd) of 95% CIs
Fisher's Z method	0.948	0.484 (0.140)
Bootstrap BC$_a$	0.936	0.464 (0.156)
Profile likelihood	0.918	0.436 (0.131)
Bayes (uniform prior)	0.964	0.522 (0.128)
Combination	0.954	0.226 (0.041)

coverage and average length of 95% CIs. We see that not only the combined approach maintains the desired coverage but also the length of CIs is roughly half of the lengths of CIs from individual studies. This result is as expected, since theoretically each study provides a $n^{-1/2}$-CIs and the sample size of combined data is $4n$ so we expect to obtain $(4n)^{-1/2}$-CI.

Data [23]:
Baseline: 2.12, 4.35, 3.39, 2.51, 4.04, 5.10, 3.77, 3.35, 4.10, 3.35, 4.15, 3.56, 3.39, 1.88, 2.56, 2.96, 2.49, 3.03, 2.66, 3.00
One year: 2.47, 4.61, 5.26, 3.02, 6.36, 5.93, 3.93, 4.09, 4.88, 3.81, 4.74, 3.29, 5.55, 2.82, 4.23, 3.23, 2.56, 4.31, 4.37, 2.40

References

1. Baath, R.: Bayesian First Aid. R package (2013)
2. Bayarri, M.J., Berger, J.O., Forte, A., García-Donato, G.: Criteria for Bayesian model choice with application to variable selection. Ann. Stat. **40**, 1550–1577 (2012)
3. Beaumont, M.A., Zhang, W., Balding, D.J.: Approximate Bayesian computation in population genetics. Genetics **162**, 2025–2035 (2002)
4. Beran, R.: Balanced simultaneous confidence sets. J. Am. Stat. Assoc. **83**, 679–686 (1988)
5. Berger, J.: Catalog of Objective Priors. Tech. rep., Duke University (2011)
6. Berger, J.O., Pericchi, L.R.: Objective Bayesian methods for model selection: introduction and comparison. In: Model Selection, vol. 38 of IMS Lecture Notes Monogr. Ser., pp. 135–207. Inst. Math. Statist., Beachwood, OH (2001)
7. Berger, J.O., Bernardo, J.M., Sun, D.: The formal definition of reference priors. Ann. Stat. **37**, 905–938 (2009)
8. Berger, J.O., Bernardo, J.M., Sun, D.: Objective priors for discrete parameter spaces. J. Am. Stat. Assoc. **107**, 636–648 (2012)
9. Birnbaum, A.: Confidence curves: An omnibus technique for estimation and testing statistical hypotheses. J. Am. Stat. Assoc. **56**, 246–249 (1961)
10. Canty, A., Ripley, B.D.: boot: Bootstrap R (S-Plus) Functions. R package version 1.3-23 (2019)
11. Casella, G., Berger, R.L.: Statistical Inference, 2nd edn. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, CA (2002)
12. Chen, X., Xie, M.: A split-and-conquer approach for analysis of extraordinarily large data. Statistica Sinica **24**, 1655–1684 (2014)
13. Cheng, J.Q., Liu, R.Y., Xie, M.: Fusion Learning, pp. 1–8 (2017). https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat07922

14. Cisewski, J., Hannig, J.: Generalized fiducial inference for normal linear mixed models. Ann. Stat. **40**, 2102–2127 (2012)
15. Claggett, B., Xie, M., Tian, L.: Meta-analysis with fixed, unknown, study-specific parameters. J. Am. Stat. Assoc. **109**, 1660–1671 (2014)
16. Cox, D.R.: Some problems connected with statistical inference. Ann. Math. Stat. **29**, 357–372 (1958). https://doi.org/10.1214/aoms/1177706618.
17. Cui, Y., Hannig, J.: Nonparametric generalized fiducial inference for survival functions under censoring (with discussions and rejoinder). Biometrika **106**, 501–518 (2019a). https://doi.org/10.1093/biomet/asz016
18. Cui, Y., Hannig, J.: Rejoinder: Nonparametric generalized fiducial inference for survival functions under censoring. Biometrika **106**, 527–531 (2019b). https://doi.org/10.1093/biomet/asz032
19. Cui, Y., Hannig, J.: A fiducial approach to nonparametric deconvolution problem: discrete case. Science China Mathematics. In press
20. Cui, Y., Hannig, J., Kosorok, M.: A unified nonparametric fiducial approach to interval-censored data. (2021) arXiv:2111.14061
21. Davison, A.C., Hinkley, D.V.: Bootstrap Methods and Their Applications. Cambridge University Press, Cambridge (1997). http://statwww.epfl.ch/davison/BMA/. ISBN 0-521-57391-2
22. Dempster, A.P.: The Dempster-Shafer calculus for statisticians. Int. J. Approx. Reason. **48**, 365–377 (2008)
23. DiCiccio, T.J., Efron, B.: Bootstrap confidence intervals. Stat. Sci. **11**, 189–228 (1996). https://doi.org/10.1214/ss/1032280214
24. Edlefsen, P.T., Liu, C., Dempster, A.P.: Estimating limits from Poisson counting data using Dempster–Shafer analysis. Ann. Appl. Stat. **3**, 764–790 (2009)
25. Efron, B.: R.A. Fisher in the 21st century. Stat. Sci. **13**, 95–122 (1998)
26. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. No. 57 in Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton, FL, USA (1993)
27. Fisher, R.A.: Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika **10**, 507–521 (1915). http://www.jstor.org/stable/2331838
28. Fisher, R.A.: Inverse probability. Proc. Camb. Phil. Soc. **xxvi**, 528–535 (1930)
29. Fraser, D.A.S.: Statistical inference: Likelihood to significance. J. Am. Stat. Assoc. **86**, 258–265 (1991)
30. Fraser, D.A.S.: Ancillaries and conditional inference. Stat. Sci. **19**, 333–369 (2004)
31. Fraser, D.A.S.: Is Bayes posterior just quick and dirty confidence? Stat. Sci. **26**, 299–316 (2011)
32. Fraser, D.A.S., McDunnough, P.: Further remarks on asymptotic normality of likelihood and conditional analyses. Can. J. Stat. [La Revue Canadienne de Statistique] **12**, 183–190 (1984). http://www.jstor.org/stable/3314746.
33. Fraser, D., Naderi, A.: Exponential models: Approximations for probabilities. Biometrika **94**, 1–9 (2008)
34. Fraser, D., Reid, N., Wong, A.: What a model with data says about theta. Int. J. Stat. Sci. **3**, 163–178 (2005)
35. Fraser, A.M., Fraser, D.A.S., Staicu, A.-M.: The second order ancillary: A differential view with continuity. Bernoulli Off. J. Bernoulli Soc. Math. Stat. Probab. **16**, 1208–1223 (2009)
36. Fraser, D.A.S., Reid, N., Marras, E., Yi, G.Y.: Default Priors for Bayesian and frequentist inference. J. R. Stat. Soc. Ser. B **72**, 631–654 (2010)
37. Hannig, J.: On generalized fiducial inference. Statistica Sinica **19**, 491–544 (2009)
38. Hannig, J.: Discussion of "On the Birnbaum Argument for the Strong Likelihood Principle" by D. G. Mayo. Stat. Sci. **29**, 254–258 (2014)
39. Hannig, J., Lee, T.C.M.: Generalized fiducial inference for wavelet regression. Biometrika **96**, 847–860 (2009)
40. Hannig, J., Xie, M.: A note on Dempster-Shafer recombinations of confidence distributions. Electr. J. Stat. **6**, 1943–1966 (2012)
41. Hannig, J., Iyer, H., Lai, R.C., Lee, T.C.: Generalized fiducial inference: A review and new results. J. Am. Stat. Assoc. **111**, 1346–1361 (2016)
42. Iverson, T.: Generalized fiducial inference. Wiley Interdiscip. Rev. Comput. Stat. **6**, 132–143 (2014) https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1291
43. Lai, R.C.S., Hannig, J., Lee, T.C.M.: Generalized fiducial inference for ultra-high dimensional regression. J. Am. Stat. Assoc. **110**, 760–772 (2015a)
44. Lai, R.C.S., Hannig, J., Lee, T.C.M.: Generalized fiducial inference for ultrahigh-dimensional regression. J. Am. Stat. Assoc. **110**, 760–772 (2015b)
45. Lawless, J.F.: Statistical Models and Methods for Lifetime Data. Wiley, New York (1982)
46. Lawless, J.F., Fredette, M.: Frequentist prediction intervals and predictive distributions. Biometrika **92**, 529–542 (2005). https://doi.org/10.1093/biomet/92.3.529
47. Lehmann, E.L.: The fisher, neyman-pearson theories of testing hypotheses: One theory or two? J. Am. Stat. Assoc. **88**, 1242–1249 (1993). http://www.jstor.org/stable/2291263.
48. Li, Y., Gillespie, B.W., Shedden, K., Gillespie, J.A.: Profile likelihood estimation of the correlation coefficient in the presence of left, right or interval censoring and missing data. R J. **10**, 159–179 (2018). https://doi.org/10.32614/RJ-2018-040
49. Liu, Y., Hannig, J.: Generalized fiducial inference for binary logistic item response models. Psychometrika **81**, 290–324 (2016)
50. Liu, Y., Hannig, J.: Generalized fiducial inference for logistic graded response models. Psychometrika **82**, 1097–1125 (2017)
51. Liu, R.Y., Singh, K.: Notions of limiting p values based on data depth and bootstrap. J. Am. Stat. Assoc. **92**, 266–277 (1997)
52. Liu, D., Liu, R.Y., Xie, M.: Exact meta-analysis approach for discrete data and its application to 2×2 tables with rare events. J. Am. Stat. Assoc. **109**, 1450–1465 (2014)
53. Liu, D., Liu, R.Y., Xie, M.: Multivariate meta-analysis of heterogeneous studies using only summary statistics: Efficiency and robustness. J. Am. Stat. Assoc. **110**(509), 326–340 (2015)
54. Luo, X., Dasgupta, T., Xie, M., Liu, R.: Leveraging the fisher randomization test using confidence distributions: inference, combination and fusion learning. Preprint (2020). arXiv:2004.08472
55. Marden, J.I.: Sensitive and sturdy p-values. Ann. Stat. **19**, 918–934 (1991). http://www.jstor.org/stable/2242091
56. Martin, R.: Inferential Models, pp. 1–8. American Cancer Society (2017). https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat07997
57. Martin, R.: On an inferential model construction using generalized associations. J. Stat. Plann. Infer. **195**, 105–115 (2018). http://www.sciencedirect.com/science/article/pii/S0378375816301537. Confidence distributions
58. Martin, R.: Discussion of Nonparametric generalized fiducial inference for survival functions under censoring. Biometrika **106**, 519–522 (2019). https://doi.org/10.1093/biomet/asz022
59. Martin, R., Liu, C.: Inferential models: A framework for prior-free posterior probabilistic inference. J. Am. Stat. Assoc. **108**, 301–313 (2013)
60. Martin, R., Liu, C.: Conditional inferential models: combining information for prior-free probabilistic inference. J. R. Stat. Soc., Ser. B **77**, 195–217 (2015a)

61. Martin, R., Liu, C.: Inferential Models: Reasoning with Uncertainty. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press (2015b). https://books.google.com/books?id=OdSYCgAAQBAJ

62. Martin, R., Liu, C.: Marginal inferential models: prior-free probabilistic inference on interest parameters. J. Am. Stat. Assoc. **110**, 1621–1631 (2015c)

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

64. Nair, V.N.: Confidence bands for survival functions with censored data: a comparative study. Technometrics **26**, 265–275 (1984)

65. Neupert, S.D., Hannig, J.: BFF: Bayesian, fiducial, frequentist analysis of age effects in daily diary data. J. Gerontol. Ser. B (2019). https://doi.org/10.1093/geronb/gbz100. Gbz100

66. Neyman, J.: Fiducial argument and the theory of confidence intervals. Biometrika **32**, 128–150 (1941). http://www.jstor.org/stable/2332207

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

68. Qiu, Y., Zhang, L., Liu, C.: Exact and efficient inference for partial bayes problems. Electron. J. Stat. **12**, 4640–4668 (2018). https://doi.org/10.1214/18-EJS1511

69. Roy, A., Mathew, T.: A generalized confidence limit for the reliability function of a two-parameter exponential distribution. J. Stat. Plann. Infer. **128**, 509–517 (2005)

70. Schein, P.S.: A comparison of combination chemotherapy and combined modality therapy for locally advanced gastric carcinoma. Cancer **49**, 1771–1777 (1982)

71. Schweder, T.: Confidence nets for curves. Adv. Stat. Model. Infer. Essays Honor Kjell A. Doksum, 593–609 (2007)

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

73. Schweder, T., Hjort, N.L.: Frequentist analogues of priors and posteriors. Econ. Philos. Econ., 285–317 (2003)

74. Schweder, T., Hjort, N.L.: Confidence, Likelihood, Probability, vol. 41. Cambridge University Press (2016)

75. Shen, J., Liu, R.Y., Xie, M.: ifusion: Individualized fusion learning. J. Am. Stat. Assoc. **0**, 1–17 (2019)

76. Signorell, A., et al.: DescTools: Tools for Descriptive Statistics. https://cran.r-project.org/package=DescTools. R package version 0.99.28 (2019)

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

78. Singh, K., Xie, M., Strawderman, W.E.: Confidence distribution (cd): Distribution estimator of a parameter. Lect. Notes Monogr. Ser. **54**, 132–150 (2007). http://www.jstor.org/stable/20461464

79. Taraldsen, G., Lindqvist, B.H.: Fiducial theory and optimal inference. Ann. Stat. **41**, 323–341 (2013)

80. Taraldsen, G., Lindqvist, B.H.: Discussion of Nonparametric generalized fiducial inference for survival functions under censoring. Biometrika **106**, 523–526 (2019). https://doi.org/10.1093/biomet/asz027

81. Tian, L., Cai, T., Pfeffer, M.A., Piankov, N., Cremieux, P.-Y., Wei, L.: Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2×2 tables with all available data but without artificial continuity correction. Biostatistics **10**, 275–281 (2008)

82. Tian, L., Wang, R., Cai, T., Wei, L.-J.: The highest confidence density region and its usage for joint inferences about constrained parameters. Biometrics **67**, 604–10 (2011)

83. Tsui, K.-W., Weerahandi, S.: Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. J. Am. Stat. Assoc. **84**, 602–607 (1989)

84. Tsui, K.-W., Weerahandi, S.: Corrections: Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. J. Am. Stat. Assoc. **84**(406), 602–607 (1991); MR1010352 (90g:62047); J. Am. Stat. Assoc. **86**, 256 (1989)

85. Veronese, P., Melilli, E.: Fiducial and Confidence Distributions for Real Exponential Families. Scand. J. Stat. (2014, in press)

86. Wandler, D.V., Hannig, J.: Generalized fiducial confidence intervals for extremes. Extremes **15**, 67–87 (2012)

87. Wang, Y.H.: Fiducial intervals: what are they? Am. Stat. **54**, 105–111 (2000)

88. Williams, J.P., Hannig, J.: Nonpenalized variable selection in high-dimensional linear model settings via generalized fiducial inference. Ann. Stat. **47**, 1723–1753 (2019). https://doi.org/10.1214/18-AOS1733

89. Williams, J.P., Storlie, C.B., Therneau, T.M., Jr., C.R.J., Hannig, J.: A bayesian approach to multistate hidden markov models: Application to dementia progression. J. Am. Stat. Assoc. **0**, 1–21 (2019)

90. Xie, M., Singh, K.: Confidence distribution, the frequentist distribution estimator of a parameter: A review. Int. Stat. Rev. **81**, 3–39 (2013)

91. Xie, M., Singh, K., Strawderman, W.E.: Confidence distributions and a unified framework for meta-analysis. J. Am. Stat. Assoc. **106**, 320–333 (2011)

92. Xie, M., Liu, R.Y., Damaraju, C.V., Olson, W.H.: Incorporating external information in analyses of clinical trials with binary outcomes. Ann. Appl. Stat. **7**, 342–368 (2013)

93. Xu, X., Li, G.: Fiducial inference in the pivotal family of distributions. Sci. China Ser. A Math. **49**, 410–432 (2006)

94. Yang, G., Liu, D., Wang, J., Xie, M.: Meta-analysis framework for exact inferences with application to the analysis of rare events. Biometrics **72**, 1378–1386 (2016). https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12497

95. Yang, G., Cheng, J.Q., Xie, M., Qian, W.: gmeta: Meta-Analysis via a Unified Framework of Confidence Distribution. https://CRAN.R-project.org/package=gmeta. R package version 2.3-0 (2017)

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

Yifan Cui received his Ph.D. from University of North Carolina at Chapel Hill in 2018. He is currently a tenure track faculty member at the Center for Data Science in Zhejiang University (ZJU). He has worked at Wharton Statistics Department, University of Pennsylvania as a postdoctoral researcher and Department of Statistics and Data Science, National University of Singapore as an assistant professor. His main research interests include nonparametric and semiparametric statistics, random forests, causal inference, precision medicine, survival analysis, generalized fiducial inference, and foundations of statistics.

Min-ge Xie receives his Ph.D. from University of Illinois at Urbana-Champaign in 1996. He is Distinguished Professor and Director of the Office of Statistical Consulting in Department of Statistics, Rutgers, The State University of New Jersey. His main research interest is in the foundation of statistical inference and fusion learning. His other expertise includes estimating equations, hierarchical models, asymptotics and applications in biostatistics and industry.