Statistica Sinica **19** (2009), 491-544

ON GENERALIZED FIDUCIAL INFERENCE*

Jan Hannig

The University of North Carolina at Chapel Hill

Abstract: In this paper we extend Fisher's fiducial argument and obtain a generalized fiducial recipe that greatly expands the applicability of fiducial ideas. We do this assuming as little structure as possible. We demonstrate this recipe on many examples of varying complexity. We investigate, by simulation and by theoretical considerations, some properties of the statistical procedures derived by the generalized fiducial recipe observing very good performance. We also compare the properties of generalized fiducial inference to the properties of Bayesian inference.

Key words and phrases: Asymptotics, fiducial inference, generalized inference, MCMC, mixture of normal distributions, multinomial distribution, structural inference.

1. Introduction

R. A. Fisher's fiducial inference has been the subject of many discussions and controversies ever since he introduced the idea during the 1930's. The idea experienced a bumpy ride, to say the least, during its early years and one can safely say that it eventually fell into disfavor among mainstream statisticians. However, it appears to have made a resurgence recently under the label of generalized inference. In this new guise, fiducial inference has proved to be a useful tool for deriving statistical procedures for problems where frequentist methods with good properties were previously unavailable. Therefore we believe that the fiducial argument of R. A. Fisher deserves a fresh look from a new angle.

Our main goal is to show that the idea of transferring randomness from the model to the parameter space seems to be a useful one—giving us a tool to design useful statistical methods. We depart from the usual tradition in several ways. When defining fiducial distribution we do not start with a pivotal quantity. Instead we start with a data generating equation also called a structural equation. This often makes no difference to the final result but it gives us the added flexibility of being able to treat continuous and discrete data in a unified fashion. We then approach the definition of a fiducial probability as a simple transfer of probability measure. We then investigate some particular examples and notice that statistical methods designed using the fiducial reasoning have typically very

*This is an exploratory paper that received special approval for exceeding the normal page limit set by Statistica Sinica.

good statistical properties as measured by their repeated sampling (frequentist) performance. This is demonstrated both in simulations and by some asymptotic considerations.

Thus fiducial inference can be viewed as a procedure that obtains a measure on a parameter space while assuming less than Bayesian inference does (no prior); it can also be viewed, as shown by our asymptotic results, as a procedure that in a routine algorithmic way defines approximate pivots for parameters of interest, which is one of the main goals of frequentist inference. Moreover, our research shows that fiducial distributions can be related to empirical Bayes methods.

Unfortunately, we also demonstrate that there is typically no unique way to define a fiducial distribution. One aspect of the non-uniqueness is related to problems associated with conditioning on an event of probability zero known as the Borel paradox—see Casella and Berger (2002, Sec. 4.9.3). Fortunately, when the model has complete sufficient statistics, which is true in many practical situations, one gets essentially unique procedures. Moreover, even in the cases when the minimal sufficient statistics is not complete we offer a particular way of deriving a fiducial distribution that works very well in a wide range of problems.

We do not attempt to derive a new "paradox free theory of fiducial inference" as we do not believe this is possible. Instead we assume as little structure as possible, present a simple recipe that can be used regardless of the dimension of the parameter space and that is easily implementable in practical applications, and we study properties of the procedures it produces.

This discussion should be also of interest to people using generalized inference procedures. The reason is that any fiducial distribution can be understood as a distribution on the parameter space implied by a particular generalized pivot and most, if not all, generalized pivotal inference procedures in the published literature are identical to procedures obtained using fiducial inference Hannig, Iyer and Patterson (2006b). In fact our generalized fiducial recipe has been developed as a generalization of the idea of a generalized pivot. Therefore, most ideas presented here are directly applicable for generalized inference as well.

The rest of this paper is organized as follows. In Section 2 we briefly discuss some aspects of the history of the fiducial argument. Section 3 gives a heuristic explanation of the fiducial argument. This is followed in Section 4 by a technical formulation of a generalized fiducial recipe, suggestion of a particular implementation of the generalized fiducial recipe, and a discussion of practical uses of the fiducial distribution together with its connection to Bayesian inference. Section 5 gives some sufficient conditions under which generalized fiducial distribution leads to confidence sets that are asymptotically exact in the frequentist sense. Section 6 gives several examples that are of independent interest. In particular we derive a generalized fiducial distribution for a variance component problem, a multinomial distribution, and a mixture of two normal populations. In each case we evaluate the performance of the proposed method by simulation. Finally, in Section 7 we give examples of non-uniqueness of generalized fiducial distribution.

2. History

R. A. Fisher introduced the idea of fiducial probability and fiducial inference Fisher (1930) in an attempt to overcome what he saw as a serious deficiency of the Bayesian approach to inference-use of a prior distribution on model parameters even when no information was available regarding their values. Although he discussed fiducial inference in several subsequent papers, there appears to be no rigorous definition of a fiducial distribution for a vector parameter θ based on sample observations. In the case of a one-parameter family of distributions, Fisher gave the following definition for a fiducial density $f(\theta|x)$ of the parameter based on a single observation x for the case where the cdf $F(x|\theta)$ is a monotonic decreasing function of θ:

$$f(\theta|x) = -\frac{\partial F(x|\theta)}{\partial \theta}. \tag{2.1}$$

Fisher illustrated the application of fiducial probabilities by means of a numerical example consisting of four pairs of observations from a bivariate normal distribution with unknown mean vector and covariance matrix. For this example he derived fiducial limits (one-sided interval estimates) for the population correlation coefficient ρ. Fisher proceeded to refine the concept of fiducial inference in several subsequent papers (Fisher (1933, 1935a)). In his 1935 paper titled "The Fiducial Argument in Statistical Inference" Fisher explained the notion of fiducial inference for μ based on a random sample from a $N(\mu, \sigma^2)$ distribution where σ is unknown. The process of obtaining a fiducial distribution for μ was based on the availability of the student's t-statistic that served as a pivotal quantity for μ. In this same 1935 paper, Fisher discussed the notion of a fiducial distribution for a single *future* observation x from the same $N(\mu, \sigma^2)$ distribution based on a random sample $x_1, \ldots, x_n$. For this he used the fact that

$$T = \frac{x - \bar{x}}{s/\sqrt{n}}$$

is a pivotal quantity. He then proceeded to consider the fiducial distribution for $\bar{x}'$ and s', the mean and the standard deviation, respectively, of m future observations $x_{n+1}, \ldots, x_{n+m}$. By letting m tend to infinity, he obtained a *simultaneous fiducial distribution* for μ and σ. He also stated "In general, it appears that if statistics $T_1, T_2, \ldots$ contain jointly the whole of the information available respecting parameters $\theta_1, \theta_2, \ldots$, and if functions $t_1, t_2, \ldots$ of the T's and θ's can be found, the simultaneous distribution of which is independent of $\theta_1, \theta_2, \ldots$, then the fiducial distribution of $\theta_1, \theta_2, \ldots$ simultaneously may be found by substitution." In essence Fisher had proposed a recipe for constructing simultaneous fiducial distributions for vector parameters. He applied this recipe to the problem of interval estimation of $\mu_1 - \mu_2$ based on independent samples from two normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with unknown means and variances, the celebrated Behrens-Fisher problem. Fisher noted that the resulting inference regarding $\mu_1 - \mu_2$ coincided with the approach proposed much earlier by Behrens (1929). He alluded to the test of the null hypothesis of *no difference*, based on the fiducial distribution of $\mu_1 - \mu_2$ as an *exact test*. This resulted in much controversy as it was noted by Fisher's contemporaries that the Behrens-Fisher test was not an exact test in the usual frequentist sense. Moreover, this same test had been obtained by Jeffreys (1940) using a Bayesian argument with non-informative priors (now known as Jeffreys priors). Fisher argued that, while Jeffreys approach gave the same answer as the fiducial approach, the logic behind Jeffreys derivation was unacceptable because of the use of an unjustified prior distribution on the parameters. Fisher particularly objected to the practice of using uniform priors to model ignorance. This led to further controversy especially between Fisher and Jeffreys.

In the same 1935 paper, Fisher gave a second example of application of his recipe by deriving a fiducial distribution for ϕ in the balanced one-way random effects model

$$Y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, n_1;\ j = 1, \ldots, n_2,$$

where $a_i \sim N(0, \phi)$, $e_{ij} \sim N(0, \theta)$, and all random variables are independent. An issue that arose from his treatment of this problem is that the fiducial distribution assigned a positive probability to the event $\phi < 0$ in spite of the fact that ϕ is a variance. Recently, using the ideas from this paper E, Hannig and Iyer (2008) provided a fiducial solution for the unbalanced version of the one-way random effects model. It is interesting to remark that simulations reported in E, Hannig and Iyer (2008) suggest that the fiducial solution leads to confidence intervals for ϕ that are conservative, but have an expected length that is shorter than all other frequentist solutions available in the literature.

Fisher's 1935 paper resulted in a flurry of activity in fiducial inference. Most of this activity was directed toward finding deficiencies in fiducial inference and philosophical concerns regarding the interpretation of fiducial probability. The controversy seems to have risen once Fisher's contemporaries realized that, unlike the case in early simple applications involving a single parameter, fiducial inference often led to procedures that were not exact in the frequentist sense. For a detailed discussion of the controversies concerning fiducial inference, the reader is referred to Zabell (1992). On a positive note Fraser, in a series of articles Fraser (1961, 1966) and a monograph Fraser (1968), attempted to provide a rigorous framework for making inferences along the lines of Fisher's fiducial inference. Fraser proposed to resolve the problems of non-uniqueness by assuming that the statistical model was coupled with an additional group structure, e.g., location-scale model. He termed his approach *structural inference*. While the presence of this additional group structure could be used as a guidance tool in resolving some sources of non-uniqueness, other sources of non-unqueness still remain; see Remark 7 in Section 7 for more details. Fraser also introduced the concept of modeling the data as a function of the parameters and some error random variable through a structural equation. This is in contrast to the more usual pivotal equation approach that gets the error random variable as a function of data and the parameter.

Additional important references include Wilkinson (1977), who attempted to explain and/or resolve some of the controversies regarding fiducial inference, Dawid and Stone (1982), who provided further insight by, among other things, studying situations where fiducial inference led to exact confidence statements, and more recently, Barnard (1995), who proposed a view of fiducial distribution based on the pivotal approach that does not seem to suffer some of the problems reported in earlier literature. However, Barnard (1995) achieves this by ignoring some of the information available in the data and restricting sets on which fiducial distribution can be evaluated. A reader interested in additional references on fiducial inference can consult Salome (1998). Nevertheless, it is fair to say that fiducial inference failed to secure a place in mainstream statistics.

In Tsui and Weerahandi (1989), a new approach was proposed for constructing hypothesis tests using the concept of *generalized P values* and this idea was later extended to a method of constructing *generalized confidence intervals* using *generalized pivotal quantities (GPQs)* Weerahandi (1993). Several papers have appeared since, in leading statistical journals, where confidence intervals have been constructed using generalized pivotal quantities in problems where exact frequentist solutions are unavailable. For a thorough exposition of generalized inference see Weerahandi (2004). Iyer and Patterson (2002) and Hannig, Iyer and Patterson (2006b) noted that every published generalized confidence interval was obtainable using the fiducial/structural arguments. In fact, Hannig, Iyer and Patterson (2006b) not only established a clear connection between fiducial intervals and generalized confidence intervals, but also proved the asymptotic frequentist correctness of such intervals. They further provided some general methods for constructing GPQs. In particular, they showed that a special class of GPQs, called fiducial GPQs (FGPQ), provide a direct frequentist interpretation to fiducial inference. However, all these articles focus on continuous distributions and do not address discrete distributions.

There is very little published literature dealing with fiducial inference for parameters of a discrete distribution. Even for the single parameter case such as the binomial distribution Fisher was aware that there were difficulties with defining a unique fiducial density for the unknown binomial parameter p. In his 1935 paper (Fisher (1935b)) titled "The Logic of Inductive Inference", Fisher gives an example where he suggests a clever device for "*turning a discontinuous distribution, leading to statements of fiducial inequality, into a continuous distribution, capable of yielding exact fiducial statements, by means of a modification of experimental procedure.*" His device was to introduce randomization into the experimental procedure and is akin to randomized decision procedures. Inspired by Fisher's example, Stevens (1950) gave a more formal treatment of this problem where he used a supplementary random variable in an attempt to define a *unique* fiducial density for a parameter of a discrete distribution. He discussed his approach in great detail using the binomial distribution as an illustration. Unfortunately, this idea seems to have gotten lost, and subsequent researchers mostly focused on fiducial inference for continuous distributions. A notable exception is Dempster (1966, 1968) whose theory of upper and lower fiducial probabilities was designed specifically for discrete distributions. In 1996, in his Fisher Memorial Lecture at the American Statistical Association annual meetings, Efron gave a brief discussion of fiducial inference with the backdrop of the binomial distribution. He said, "*Fisher was uncomfortable applying fiducial arguments to discrete distributions because of the ad hoc continuity corrections required, but the difficulties caused are more theoretical than practical.*" See Efron (1998). In fact, Efron's suggestion for how to handle discrete distributions is a special case of Stevens (1950).

Dempster's idea of upper and lower probabilities was further developed in the Dempster-Shafer calculus (Dempster (2008)), a mathematical theory of evidence. To explain the main paradigm of this theory applied in our context, consider the following simple example. Let $X = I_{(0,p)}(U)$, where $p \in (0,1)$ is an unknown fixed number. If we found $X = 1, U = 0.3$, we could conclude that $p \in (0.3,1)$, e.g., we would know that statement $\{p < 0.1\}$ is not true, statement $\{p > 0.2\}$ is true, and $\{p > 0.9\}$ is unsure. Now more realistically, let us assume that $X = 1$ and U is an unknown realization of a $U(0,1)$ random variable. Just as before we know $p \in (U,1)$, which now is a random statement. This statement can interpreted as follows: The event $\{p < 0.1\}$ is not possible if $U > 0.1$ as the interval $(U,1)$ has empty intersection with $(0,0.1)$. Hence we assign the probability 0.9 to the statement "it is not true that$\{p < 0.1\}$". Similarly, if $U < 0.1$ the interval $(U,1)$ has non-empty intersection with both $(0,0.1)$ and its complement and therefore $\{p < 0.1\}$ is unsure with probability 0.1. In other words we assign probability of 0.1 to the statement "we do not know if $p < 0.1$". Finally, $\{p < 0.1\}$ is certain only if $U = 0$ which has probability 0. Thus we assign probability 0 to the statement "we are convinced $\{p < 0.1\}$". Similarly, $\{p > 0.7\}$ is certain with probability 0.3, because if $U > 0.7$ the interval $(U, 1)$ is included in the interval $(0.7, 1)$. The statement $\{p > 0.7\}$ is unsure with probability 0.7, because again if $U < 0.7$ the interval $(U, 1)$ has non-empty intersection with both $(0.7, 1)$ and its complement. Finally, $\{p > 0.7\}$ can be excluded only if $U = 1$ which has probability 0.

Using a more statistical terminology, the information on the parameter p is not summarized in terms of measure on the parameter space $(0, 1)$, but rather in terms of a measure on the space of subsets of the parameter space together with a rule on how to interpret this measure in terms of the parameter. The second part of Dempster-Shafer calculus is a rule on how to combine information from two such measures under the assumption of independence. This rule is too complicated to spell out here, but we remark that it bears similarities to our fiducial recipe introduced in Section 4. In particular, just as is the case with our procedure, Dempster's recombination rule suffers from non-uniqueness due to conditioning on events of probability zero—the Borel paradox. A reader interested in a more thorough introduction to Dempster-Shafer calculus for statisticians is referred to Dempster (2008).

As mentioned earlier, fiducial inference has recently made a comeback in applied literature partially under the guise of generalized inference. In the field of Metrology there is a movement to establish fiducial and generalized inference as one of the mainstream methods of that discipline, Wang and Iyer (2005, 2006a,b). Several researchers have worked on various measures of process repeatability and reproducibility using generalized confidence intervals derived from generalized pivotal quantities. See, for instance, Daniels, Burdick and Quiroz (2005), Burdick, Park, Montgomery and Borror (2005b), Hamada and Weerahandi (2000) and Iyer, Wang and Mathew (2004). Finally, McNally, Iyer and Mathew (2003) have applied the method of generalized confidence intervals to selected applications in pharmaceutical statistics. Given this flurry of recent activity generated by applications, we believe that it is important to further develop the understanding of fiducial inference and its performance in specific problems.

In this paper we provide a general definition for fiducial distributions for parameters that applies equally well to continuous as well as discrete parent distributions. The resulting inference is termed *generalized fiducial inference*, rather than *fiducial inference*, to emphasize connection with generalized inference, as well as the fact that multiple generalized fiducial distributions can be defined for the same parameter.

We close this section with some quotes. Zabell (1992) begins his Statistical Science paper with the statement "*Fiducial inference stands as R. A. Fisher's* *one great failure."* On the other hand, Efron, in his 1998 Statistical Science paper (based on his Fisher Memorial Lecture of 1996), in the section dealing with fiducial inference, has said "*I am going to begin with the fiducial distribution, generally considered to be Fisher's biggest blunder.*" However, in the closing paragraph of the same section (Section 8), he says "*Maybe Fisher's biggest blunder will become a big hit in the* 21*st century*!"

3. The Fiducial Argument

The main aim of fiducial inference is to define a distribution for parameters of interest that captures all of the information that the data contains about these parameters. This *fiducial distribution* can later be used for proposing inference procedures such as confidence sets. In this sense, a fiducial distribution is much like a Bayesian posterior distribution. Fisher wanted to accomplish this without assuming a prior distribution on the parameters.

We would like to introduce our understanding of the fiducial argument by comparing it to the widely accepted notion of likelihood function. Recall that the likelihood function is obtained by considering the density function of our data, $f(\mathbf{x}, \xi)$, and switching the role of the variable and the parameter. With ξ known and fixed, the density determines the probability of observing any given value of $\mathbf{x}$, while the likelihood function considers $\mathbf{x}$ fixed and calibrates our belief in various values of ξ.

The fiducial distribution is based on a similar idea of switching the role of the parameter and the data. We start with a structural equation $X = G(\xi, U)$ where ξ is a parameter and U is a random vector with completely known distribution independent of any parameters. Often one can think of the structural equation as a detailed description of the noise process U that combines with the signal ξ to yield observed data X. Thus for any fixed value of the parameter ξ the distribution of U and the structural equation imply the distribution of the data X. After observing the data X we can switch the role of data and parameters. In particular, we fix the value of X and use the distribution of U and the structural equation (this time considered as an implicit equation) to infer a distribution on ξ. In other words, one can get a random realization from the fiducial distribution of ξ by generating U and solving the structural equation for ξ, conditioning on the fact that the solution exists. We now proceed to demonstrate this idea on a simple example.

Consider a random variable X from a normal distribution with unknown mean μ and variance 1, i.e., $X = \mu + Z$ where Z is standard normal. If x is a realized value of X corresponding to the realized value z of Z, then we have $\mu = x - z$. Of course the value z is not observed. However, a contemplated value μ_0 of μ corresponds to the value $x - \mu_0$ of z. Knowing that z is a realization from the $N(0,1)$ distribution, we can evaluate the likelihood of Z taking on the value $x-\mu_0$. Speaking informally, one can say that the "plausibility" of the parameter μ taking on the value μ_0 "is the same" as the plausibility of the random variable Z taking on the value $x-\mu_0$. Using this rationale, we write $\mu = x - Z$ where x is regarded as fixed but Z is still considered a $N(0,1)$ random variable. This step, namely, shifting from the true relationship $\mu = x - z$ (z unobserved) to the relationship $\mu = x - Z$, is what constitutes the *fiducial argument.* We can use the relation, $\mu = x - Z$, to define a probability distribution for μ. This distribution is called the "fiducial distribution" of μ. In particular, a random variable M carrying the fiducial probability distribution of μ can be defined based on the probabilities of observing the value of Z needed to get the desired value of μ, i.e., define M so that

$$P(M \in (a,b)) = P(x - Z \in (a,b)) = P(Z \in (x-b, x-a)).$$

For theoretical considerations, it is useful to consider a particular version of this random variable defined as $M = x - Z^\star$, where $Z^\star$ is a standard normal random variable independent of Z.

In conclusion, notice that to obtain a random variable that has a distribution described above, we had to take the structural equation $X = \mu + Z$, solve for $\mu = X - Z$ and replace Z with $Z^\star$, a random variable independent of Z having the same distribution as Z, to get $M = x - Z^\star$, where x is the observed value.

4. Generalized Fiducial Recipe

We now generalize the idea described in Section 3 to arbitrary statistical models. Our definition of generalized fiducial distribution is influenced both by generalized pivotal quantities and by Fraser's structural inference—see Appendix 3 of Dawid, Stone and Zidek (1973) for a very concise description of the structural inference idea. The main difference between Fraser's proposal and the recipe presented below is that we do not assume a group structure.

Let $\mathbb{X}$ be a (possibly discrete) random vector with a distribution indexed by a parameter $\xi \in \Xi$. Assume that the data generating mechanism for $\mathbb{X}$ could be expressed as

$$\mathbb{X} = G(\xi, U), \tag{4.1}$$

where G is a jointly measurable function and U is a random variable or vector with a completely known distribution independent of any parameters. The equation (4.1) can be understood as the equation that was used to generated the data, and we term it the *structural equation.* We define a set-valued function

$$Q(\mathbf{x}, u) = \{\xi : \mathbf{x} = G(\xi, u)\}. \tag{4.2}$$

The function $Q(\mathbb{X}, U)$ could be understood as an inverse of the function G. Here G defines u as an implicit function of ξ, and $\mathbf{x}$ is regarded as fixed. To avoid measurability problems, assume $Q(\mathbf{x}, u)$ is a measurable function of u.

We use the inverse function $Q(\mathbf{x}, u)$ to define the fiducial distribution on the parameter space. However, in some problems the inverse $Q(\mathbf{x}, u)$ could contain more than one element, in which case we need to select one of the element in $Q(\mathbf{x}, u)$ according to some, possibly random, rule. Mathematically this is achieved as follows: assume for any measurable set S, there is a random element $V(S)$ with support $\overline{S}$, where $\overline{S}$ is the closure of S. We then use the function $V(Q(\mathbf{x}, u))$ in our definition.

Finally, notice that the equation $\mathbf{x} = G(\xi, u)$ is satisfied for ξ, and u used to generate our observed data $\mathbf{x}$. In other words for this particular u, the function $Q(\mathbf{x}, u) \neq \emptyset$. Therefore, in addition to knowing the distribution of U we also know that the event $\{Q(\mathbf{x}, u) \neq \emptyset\}$ has happened and we will have to condition the distribution of U on this event.

Thus, we define a generalized fiducial distribution of ξ as the conditional distribution

$$V(Q(\mathbf{x}, U^\star)) \mid \{Q(\mathbf{x}, U^\star) \neq \emptyset\}. \tag{4.3}$$

Here $\mathbf{x}$ is the observed value of $\mathbb{X}$ and $U^\star$ is an independent copy of U.

It is useful for future considerations to denote a random element having the distribution described in (4.3) by $\mathcal{R}_\xi(\mathbf{x})$. We call this random variable a *generalized fiducial quantity* (GFQ). It is often of interest to provide inference procedures for $\theta = \pi(\xi) \in \mathbb{R}^q$ in which case we define the marginal fiducial distribution for θ as the distribution of

$$\mathcal{R}_\theta(\mathbf{x}) = \pi(\mathcal{R}_\xi(\mathbf{x})). \tag{4.4}$$

We also remark that GFQs are a generalization of fiducial generalized pivotal quantities (FGPQs) introduced by Hannig, Iyer and Patterson (2006b).

The following examples provide simple illustrations of the definition of a generalized fiducial distribution.

Example 1. Suppose $U = (E_1, E_2)$, where E_i are i.i.d. $N(0,1)$ and

$$X = (X_1, X_2) = G(\mu, U) = (\mu + E_1, \mu + E_2)$$

for some $\mu \in \mathbb{R}$. So the X_i are i.i.d. $N(\mu, 1)$. Given a realization $x = (x_1, x_2)$ of X, the set-valued function Q maps $u = (e_1, e_2) \in \mathbb{R}^2$ to a subset of $\mathbb{R}$, where

$$Q(x, u) = \begin{cases} \{x_1 - e_1\} & \text{if} \quad x_1 - x_2 = e_1 - e_2, \\ \emptyset & \text{if} \quad x_1 - x_2 \neq e_1 - e_2. \end{cases}$$

Notice that $Q(x,u)$ is either empty or is a singleton, hence V does not have to be considered here.

By definition, a generalized fiducial distribution for μ is the distribution of $x_1 - E_1^\star$ conditional on $E_1^\star - E_2^\star = x_1 - x_2$ where $U^\star = (E_1^\star, E_2^\star)$ is an independent copy of U. Hence a generalized fiducial distribution for μ is $N(\bar{x}, 1/2)$ where $\bar{x} = (x_1 + x_2)/2$.

Example 2. Suppose $U = (U_1, \ldots, U_n)$ is a vector of i.i.d. uniform $(0,1)$ random variables U_i. Let $p \in [0,1]$. Let $X = (X_1, \ldots, X_n)$ be defined by $X_i = I(U_i < p)$. So the X_i are i.i.d. Bernoulli random variables with success probability p. Suppose $x = (x_1, \ldots, x_n)$ is a realization of X. Let $s = \sum_{i=1}^n x_i$ be the observed number of 1's. The mapping $Q : [0,1]^n \to [0,1]$ is given by

$$Q(x,u) = \begin{cases} [0, u_{1:n}] & \text{if} \quad s = 0, \\ (u_{n:n}, 1] & \text{if} \quad s = n, \\ (u_{s:n}, u_{s+1:n}] & \text{if} \quad s = 1, \ldots, n-1 \\ & \quad \text{and} \quad \sum_{i=1}^n I(x_i = 1) I(u_i \le u_{s:n}) = s, \\ \emptyset & \text{otherwise.} \end{cases}$$

Here $u_{r:n}$ denotes the r^{th} order statistic among $u_1, \ldots, u_n$.

Notice that if $Q(x,u)$ is non-empty, it is an entire interval. Therefore we need to select a particular $V(\bullet)$ and this selection will have an effect on the final answer. A generalized fiducial distribution for p is given by the distribution of $V(Q(x, U^\star))$ conditional on the event $Q(x, U^\star) \neq \emptyset$ where $V(Q(x, U^\star))$ is any random variable whose support is contained in $Q(x, U^\star)$. By the exchangeability of $U_1^\star, \ldots, U_n^\star$ it follows that the stated conditional distribution of $V(Q(x, U^\star))$ is the same as the distribution of $V([0, U_{1:n}^\star])$ when $s = 0$, $V((U_{s:n}^\star, U_{s+1:n}^\star])$ for $0 < s < n$, and $V((U_{n:n}^\star, 1])$ for $s = n$. Based on simulations reported in Section 6 a good choice of $V((a,b])$ is $V((a,b]) = a$ with probability 1/2 and $V((a,b]) = b$ with probability 1/2. In this case the FQ $\mathcal{R}_p = BU_{s:n}^\star + (1-B)U_{s+1:n}^\star$ where B is a Bernoulli(1/2) random variable. Other choices of $V((a,b])$ lead to slightly different generalized fiducial distributions.

A similar answer can be also obtained by Dempster-Shafer calculus (Dempster (2008)). The main difference between our proposal and Dempster-Shafer calculus is the usage of the quantity $V(\bullet)$, i.e., the answer based on Dempster-Shafer calculus would not use V. It would be based on the distribution of the random interval $(U_{s:n}^\star, U_{s+1:n}^\star]$. The feasibility of a statement $p \in A$ would then be given as follows: $p \in A$ is true with probability $p_y = P((U_{s:n}^\star, U_{s+1:n}^\star] \subset A)$, false with $p_n = 1 - P((U_{s:n}^\star, U_{s+1:n}^\star] \subset A^{\complement})$ and undecidable with probability $1 - p_y - p_n$.

While Dempster-Schafer approach is theoretically appealing in its honest acknowledgment of the fact that there is some chance that we cannot say whether statement is true or not, we prefer in most applications to resolve this uncertainty by the choice of $V(\bullet)$. Such a solution leads to answers that are simpler to interpret and understand for most practitioners. When we return to this example in Section 6, we discuss the issue of choice of $V(\bullet)$ in more detail. In particular we propose the use of a particular $V(\bullet)$, already mentioned above, that leads to first order accurate statistical procedures. In any case we show in Section 6 that, as n grows, the fiducial distribution becomes insensitive to the choice of V.

Remark 1. As demonstrated in the examples above the recipe in (4.3) does not lead to a unique distribution. In fact, in addition to the well documented potential non-uniqueness due to the particular choice of the structural equation, there are two additional sources of non-uniqueness. As seen in Example 2, the first source of non-uniqueness is the choice of the random variable $V(Q(\mathbf{x}, u))$ if the set $Q(\mathbf{x}, u)$ has more than one element. This typically happens if we deal with discrete random variables. In this case the choice of $V(Q(\mathbf{x}, u))$ is necessarily subjective, though some choices could lead to better repeated sample performance than others.

The second source of non-uniqueness is more subtle and comes from the fact that in some situations we have $P(Q(\mathbf{x}, U^\star) \neq \emptyset) = 0$. This usually happens when dealing with continuous data. If $P(Q(\mathbf{x}, U^\star) \neq \emptyset) = 0$ the conditional distribution (4.3) needs to be interpreted as a conditional probability given a σ-algebra or, equivalently, we need to find a random object $H(U^\star)$ such that $\{Q(\mathbf{x}, U^\star) \neq \emptyset\} = \{H(U^\star) = 0\}$. In this case (4.3) could be interpreted as

$$V(Q(\mathbf{x}, U^\star)) \,|\, H(U^\star) = 0.$$

Unfortunately, if $P(H(U^\star) = 0) = 0$, different choices of H could lead to different conditional distributions. This is related to the Borel's paradox described, for example, in Casella and Berger (2002, Sec. 4.9.3).

Notice that the only additional information available to us is the fact that the value of $U^\star$ and $\mathbf{x}$ must be compatible. In other words $U^\star$ has to be such that $\mathbf{x} = G(\xi, U^\star)$ for at least one value of ξ, i.e., $Q(\mathbf{x}, U^\star) \neq \emptyset$. Therefore, any particular choice of H could be viewed as "added information". The theory of probability does not give us a reason to favor one choice of H over another. We might have some other reason, e.g., additional structure, that would guide us to a particular choice. However, there could be natural choices of H arising from different points of view. This could in turn give us more than one generalized fiducial distribution causing what may be viewed by some as a paradox. We discuss these issues in much greater detail in Section 7.

Remark 2. The choice of a particular form of the structural equation (4.1) could influence the generalized fiducial distribution. This problem is generally well-known. It is the aim of this paper to discuss the statistical challenges left after we fix a structural equation. Therefore in the remainder of this paper we regard data represented by a different structural equation as a different statistical problem even if they have the same distribution.

It is important to remark that in some practical applications the physical process by which the data was generated is known. In this case we can and should choose the structural equation to reflect this process. This eliminates the problem of non-uniqueness due to the choice of structural equation. A canonical example arises in the field of metrology, where an unknown quantity is measured using some known processes. The processes have known physical characteristics and add errors to the measured quantity in some pre-specified known fashion. The resulting measured values are expressed as an equation combining some unknown measured quantities and errors. This equation can be taken as the structural equation. In fact, the Guide to expression of Uncertainty in Measurements (GUM), an ISO document that is adhered to by national metrology laboratories of many countries, has an explicit requirement that all measured values be related to true values through a measurement equation that is really a structural equation in our terminology. For a particular example see Annex H.1 of GUM.

4.1. Suggested implementation of the recipe

As discussed in Remark 1, when implementing the fiducial recipe for continuous distribution, one needs to make some choices regarding the random vector used for conditioning. In what follows we recommend one way of resolving this problem and implementing the generalized fiducial recipe. This implementation worked well in the simulations we performed. We remark that similar ideas in a less general form can be already found in Hannig, Iyer and Patterson (2006b).

We first illustrate our approach on a simple (well-known) example and derive a joint generalized fiducial distribution for the parameters μ and σ^2 based on a random sample from $N(\mu, \sigma^2)$.

Example 3. Let $X_1, \ldots, X_n$ be i.i.d. $\mathrm{N}(\mu, \sigma^2)$. We can describe the distribution of $\mathbb{X}$ by means of the structural equations

$$X_i = \mu + \sigma Z_i,\ i = 1, \ldots, n.$$

Here the Z_i are i.i.d. standard normal random variables. We split our n equations into two groups: the first two equations are used to solve for μ, σ^2; than take the remaining $n-2$ equations, plug into them the solutions for μ, σ^2 obtained from the first two equations, and condition on them being true. More precisely,

$$Q(x_1,\dots,x_n;z_1,\dots,z_n)=\begin{cases}\left\{\left(\frac{z_1x_2-z_2x_1}{z_1-z_1},\left(\frac{x_1-x_2}{z_1-z_2}\right)^2\right)\right\} & \\ \qquad\text{if } x_l=\frac{z_ix_j-z_jx_i}{z_i-z_j}-\left|\frac{x_i-x_j}{z_i-z_j}\right|z_l,\ l=3,\dots,n; \\ \emptyset \qquad \text{otherwise.}\end{cases}$$

Defining

$$M=\frac{z_1x_2-z_2x_1}{z_1-z_2}, H=\left(\frac{x_1-x_2}{z_1-z_2}\right)^2, \text{ and } R_l=\frac{z_1x_2-z_2x_1}{z_1-z_2}-\left|\frac{x_1-x_2}{z_1-z_2}\right|z_l,$$

we can interpret the generalized fiducial distribution (4.3) as the conditional distribution of (M,H) given $\mathbf{R}=\mathbf{x}$, where $\mathbf{R}=(R_3,\dots,R_n)$ and $\mathbf{x}=(x_3,\dots,x_n)$. A simple calculation shows that the joint density of $(M,H,\mathbf{R})$ is

$$f_{M,H,\mathbf{R}}(m,h,x_3,\dots,x_n)=\frac{e^{-(\sum_{i=1}^n(m-x_i)^2)/(2h)}|x_1-x_2|}{2(2\pi)^{n/2}h^{n/2+1}}I_{(0,\infty)}(h).$$

However, there is no particular reason why we should have used the first two equations to solve for μ and σ. Therefore, given the symmetry natural to independent data, it seems more natural to assume that the two equations used to solve for μ and σ were selected at random. This leads to

$$f_{M,H,\mathbf{R}}(m,h,\mathbf{x})=\frac{e^{-(\sum_{i=1}^n(m-x_i)^2)/(2h)}\binom{n}{2}^{-1}\sum_{i<j}|x_i-x_j|}{2(2\pi)^{n/2}h^{n/2+1}}I_{(0,\infty)}(h). \qquad (4.5)$$

In any case a simple calculation shows that the fiducial density of $\mathcal{R}_{(\mu,\sigma^2)}=f_{M,H|\mathbf{R}=\mathbf{x}}(m,h)$ is

$$f_{\mathcal{R}_{(\mu,\sigma^2)}}(m,h)=\frac{e^{-(m-\bar{x}_n)^2/(2h/n)-(n-1)s_n^2/(2h)}((n-1)s_n^2)^{(n-1)/2}}{\sqrt{\pi/n}\,\Gamma\left((n-1)/2\right)2^{n/2}h^{n/2+1}}I_{(0,\infty)}(h). \quad (4.6)$$

This is the joint fiducial density proposed by Fisher (1935a), which is also the Bayesian posterior with respect to Jeffreys prior (Jeffreys (1961)). It is known that statistical methods based on (4.6) lead to exact frequentist inference Mood, Graybill and Boes (1974).

General Case

We now perform similar computations in a more general situation. Let us suppose that the parameter of interest ξ is p-dimensional. Recall the structural equation, $\mathbb{X}=G(\mathbf{U},\xi)$. Write $G=(g_1,\dots,g_n)$ so that $X_i=g_i(\mathbf{U},\xi)$ for $i=$

$1,\dots,n$. Set $\mathbb{X}_0 = (X_1,\dots,X_p)$, $\mathbb{X}_c = (X_{p+1},\dots,X_n)$, write $\mathbf{U}_0 = (U_1,\dots,U_p)$, $\mathbf{U}_c = (U_{p+1},\dots,U_n)$, and assume that the structural equation can be factorized into $\mathbf{G} = (\mathbf{G}_0,\mathbf{G}_c)$, where

$$\mathbb{X}_0 = \mathbf{G}_0(\xi,\mathbf{U}_0) \quad \text{and} \quad \mathbb{X}_c = \mathbf{G}_c(\xi,\mathbf{U}_c).$$

Additionally assume that for each fixed $\xi \in \Xi$ the functions $\mathbf{G}_0(\xi,\,\cdot\,)$ and $\mathbf{G}_c(\xi,\,\cdot\,)$ are one-to-one and differentiable. Thus

$$f_{\mathbb{X}}(\mathbf{x}|\xi) = f_{\mathbf{U}}(\mathbf{G}_0^{-1}(\mathbf{x}_0,\xi),\mathbf{G}_c^{-1}(\mathbf{x}_c,\xi))J_{\mathbf{G}_0^{-1}}(\mathbf{x}_0,\xi)J_{\mathbf{G}_c^{-1}}(\mathbf{x}_c,\xi). \tag{4.7}$$

Finally assume that, for each fixed $\mathbf{u}_0$, the mapping $\mathbf{G}_0(\,\cdot\,,\mathbf{u}_0)$ is invertible and differentiable. Denote this inverse mapping by $\mathbf{H}_\xi(\mathbf{x}_0,\mathbf{u}_0)$ and write $\mathbf{H}_c(\mathbf{x},\mathbf{e}) = \mathbf{G}_c(\mathbf{H}_\xi(\mathbf{x}_0,\mathbf{u}_0),\mathbf{u}_c)$, setting $\mathbf{H} = (\mathbf{H}_\xi,\mathbf{H}_c)$. Thus the definition of the generalized fiducial distribution (4.3) can be interpreted as the conditional distribution of

$$\mathbf{H}_\xi(\mathbf{x},\mathbf{U}^\star)\,|\,\mathbf{H}_c(\mathbf{x},\mathbf{U}^\star) = \mathbf{x}_c. \tag{4.8}$$

To derive the conditional density of (4.8), notice that if $\mathbf{x} = \mathbf{G}(\xi,\mathbf{e})$ then $\mathbf{H}_\xi(\mathbf{x}_0,\mathbf{u}_0) = \xi$ and $\mathbf{H}_c(\mathbf{x},\mathbf{e}) = \mathbf{x}_c$. Finally, for all fixed $\mathbf{x}$,

$$\mathbf{H}^{-1}((\zeta,\mathbf{s}),\mathbf{x}) = (\mathbf{G}_0^{-1}(\mathbf{x}_0,\zeta),\mathbf{G}_c^{-1}(\mathbf{s},\zeta))$$

and the Jacobian

$$J_{\mathbf{H}^{-1}}(\mathbf{x},\zeta) = J_{G_0^{-1}(\mathbf{x}_0,\,\cdot\,)}(\mathbf{x}_0,\zeta)J_{G_c^{-1}}(\mathbf{x}_c,\zeta).$$

Here $J_{G_0^{-1}(\mathbf{x}_0,\,\cdot\,)}$ is the Jacobian constructed by taking derivatives with respect to ζ. The joint density density of $\mathbf{H}(\mathbf{x},U^\star)$ at the point $(\zeta,\mathbf{x}_c)$ is

$$f_{\mathbf{H}}(\zeta,\mathbf{x}_c) = f_U(\mathbf{H}^{-1}((\zeta,\mathbf{x}_c),\mathbf{x}))J_{\mathbf{H}^{-1}}((\zeta,\mathbf{x}_c),\mathbf{x}). \tag{4.9}$$

By comparing (4.7) and (4.9) we get

$$f_{\mathbf{H}}(\xi,\mathbf{x}_c) = f_{\mathbb{X}}(\mathbf{x}|\xi)J_0(\mathbf{x}_0,\xi) \quad \text{where} \quad J_0(\mathbf{x}_0,\xi) = \left|\frac{\det\left(d\mathbf{G}_0^{-1}(\mathbf{x}_0,\xi)/(d\xi)\right)}{\det\left(d\mathbf{G}_0^{-1}(\mathbf{x}_0,\xi)/(d\mathbf{x}_0)\right)}\right|.$$

Therefore the conditional density of (4.8) is

$$r(\xi) = \frac{f_{\mathbf{X}}(\mathbf{x}|\xi)J_0(\mathbf{x}_0,\xi)}{\int_\Xi f_{\mathbf{X}}(\mathbf{x}|\xi')J_0(\mathbf{x}_0,\xi')\,d\xi'}. \tag{4.10}$$

The generalized fiducial distribution in (4.10) depends on the choice of G_0 and G_c, which is fairly arbitrary, as there is no particular reason why we should take the first p coordinates and use them for G_0. In fact such a factorization can be typically done based on any p coordinates $(i_1,\ldots,i_p)$. In this case it is reasonable to consider the p coordinates used in G_0 as selected at random. This leads to the following interpretation of the generalized fiducial distribution (4.3):

$$f_{\mathcal{R}_\xi}(\xi)=\frac{f_{\mathbf{X}}(\mathbf{x}|\xi)J(\mathbf{x},\xi)}{\int_\Xi f_{\mathbf{X}}(\mathbf{x}|\xi')J(\mathbf{x},\xi')\,d\xi'},\tag{4.11}$$

where

$$J(\mathbf{x},\xi)=\binom{n}{p}^{-1}\sum_{\mathbf{i}=(i_1,\ldots,i_p)}\left|\frac{\det\left(d\mathbf{G}_{0,\mathbf{i}}^{-1}(\mathbf{x}_\mathbf{i},\xi)/(d\xi)\right)}{\det\left(d\mathbf{G}_{0,\mathbf{i}}^{-1}(\mathbf{x}_\mathbf{i},\xi)/(d\mathbf{x}_\mathbf{i})\right)}\right|\tag{4.12}$$

is the sum taken over all subsets of indexes $1\le i_1<\cdots<i_p\le n$, and $\mathbf{G}_{0,\mathbf{i}}$ is the function $\mathbf{G}_0$ with the indexes $\mathbf{i}$ used to do the factorization.

One-parameter Case

It is of interest to show the form of of (4.11) in the case of one-parameter problems. Let $X_1,\ldots,X_n$ be independent continuous random variables each with density $f_i(x_i|\xi)$ and distribution function $F_i(x_i|\xi)$ respectively, where $\xi\in\mathbb{R}$. Choose the structural equation $X_i=F_i^{-1}(\theta,U)$ and assume that the various invertibility conditions stated above are satisfied. In light of this structural equation, (4.12) becomes

$$J_{0,i}(x_i,\theta)=\frac{|\frac{\partial}{\partial\theta}F_i(x_i,\theta)|}{f_i(x_i,\theta)}\quad\text{and}\quad J(\mathbf{x},\theta)=n^{-1}\sum_{i=1}^{n}J_{0,i}(x_i,\theta),\tag{4.13}$$

and the recommended generalized fiducial density in (4.11) is

$$f_{\mathcal{R}_\theta}(\theta)=\frac{\left(\prod_{j=1}^n f(x_i,\theta)\right)J(\mathbf{x},\theta)}{\int_\Theta\left(\prod_{j=1}^n f(x_i,\theta')\right)J(\mathbf{x},\theta')\,d\theta'}.\tag{4.14}$$

Notice that if $n=1$, the generalized fiducial density described in (4.14) is Fisher's original definition of fiducial distribution.

4.2. Practical matters and relationship with Bayesian inference

Since the distribution of $\mathcal{R}_\xi(\mathbf{x})$ for each observed $\mathbf{x}$ is known (or at least accessible through simulations), we can use it for statistical inference. The hope is that any inference procedure based on the distributions of $\mathcal{R}_\xi(\mathbf{x})$ should give reasonably good answers. To assess the quality of a statistical procedure we use the frequentist repeated sample paradigm, e.g., coverage and expected length of confidence intervals.

Since a generalized fiducial distribution provides us with a distribution on the parameter space, its use is similar to the practical use of a Bayesian posterior. For example, we can take the expected value, the median, or some other functional of the generalized fiducial distribution to get a point estimator of the parameter ξ. More importantly, we can find sets $C(\mathbf{x})$ with fiducial probability $P(\mathcal{R}_\xi(\mathbf{x}) \in C(\mathbf{x})) = 1-\alpha$ and use them as approximate $(1-\alpha)100\%$ confidence sets. These confidence sets, though not exact, seem to have very good coverage/expected length properties in small sample simulations, and often are exact asymptotically, see Section 5.

In addition to finding approximate confidence sets, which is the application addressed here, we could also use the generalized fiducial distribution for hypothesis testing and approximate p-values. In particular, define a family of regions $H(\mathbf{x},\alpha)$ such that, for each fixed $\mathbf{x}$, the sets $H(\mathbf{x},\alpha)$ are nested and $P(\mathcal{R}_\theta(\mathbf{x}) \in H(\mathbf{x},\alpha)) = 1-\alpha$. Then, for each observed $\mathbf{x}$, the fiducial p-value would be defined as $\sup_{\xi_0 \in \mathcal{H}}(\sup_{\xi_0 \in H(\mathbf{x},\alpha)} \alpha)$. The form of the regions $H(\mathbf{x},\alpha)$ is determined based on the alternative hypothesis.

Generalized fiducial distributions can also be used for prediction. This is done by combining the generalized fiducial distribution on the parameters with a structural equation for the new observations. This approach produces a predictive distribution that accommodates in a natural way both the uncertainty in the parameter estimation and the randomness of the future data.

There is another practical issue that generalized fiducial inference shares with Bayesian inference. Generalized fiducial distributions are rarely available in closed form. Therefore we often need to use an MCMC method such as a Metropolis-Hastings or Gibbs sampler to obtain a sample from the generalized fiducial distribution. While the basic issues facing implementation of the MCMC procedures are similar for both Bayesian and generalized fiducial problems, there are specific challenges related to generalized fiducial procedures.

A careful reader might ask at this point the following natural question. Given the strong similarities between the use of generalized fiducial and Bayesian inference is there any difference between the two? The answer is yes.

First, there is a basic philosophical difference. Bayesian approach starts with a fully specified, single joint probability distribution for $(\mathbf{X},\xi)$. Then it predicts a value of ξ given $\mathbf{X}$ using conceptually simple probabilistic computations (Bayes Theorem). The specification of this single probability distribution for $(\mathbf{X},\xi)$ is usually done by selecting a model $f(\mathbf{x}|\xi)$ and a prior $\pi(\xi)$, with the choice of $\pi(\xi)$ that can be viewed, in the absence of any prior information, as arbitrary. The fiducial approach is similar to the usual frequentist approach in the modeling step, as it considers a number of potential distributions for the observed data as the model. The idea is that the model is incorporating all the known prior information and assumptions. Then the fiducial distribution considers a likelihood-like idea of switching the role of data and parameters to introduce the distribution on the parameter space. This distribution then summarizes our knowledge, including uncertainty, about the unknown (fixed) parameter.

Second, generalized fiducial distribution often cannot be obtained as a Bayesian distribution with respect to any (proper or improper) prior. An early example is due to Grundy (1956). This can be understood by carefully studying the fiducial density in (4.11). This density on the parameter space is visually similar to the usual Bayes posterior with the role of a prior served by the data dependent function $J(\mathbf{x}, \xi)$. Thus the proposed generalized fiducial distribution is Bayesian posterior if and only if $J(\mathbf{x}, \xi) = k(\mathbf{x})l(\xi)$ where k and l are measurable functions. However $J(\mathbf{x}, \xi)$ typically does not decompose in this way, in which case the fiducial distribution is not a Bayesian posterior with respect to any proper or improper prior.

There is one more interesting connection to Bayesian inference. For each fixed ξ, the quantity $J(\mathbf{x}, \xi)$ in (4.12) is, by definition, a U-statistic. Therefore, if our data is i.i.d. we are guaranteed an a.s. convergence to $\pi(\xi) = E_{\xi_0} J_0(X_1, \ldots, X_p, \xi)$. In fact, under slightly stronger assumptions on the continuity of the Jacobians J_0 we obtain, using Yeo and Johnson (2001), that the convergence is uniform in ξ on compact sets. At first glance $\pi(\xi)$ could be considered as an interesting non-subjective prior. Unfortunately this prior cannot be used in applications because the expectation in the definition of $\pi(\xi)$ is taken with respect to the true unknown parameter ξ_0. However, the quantity $J(\mathbf{x}, \xi)$ could be considered as an estimator of $\pi(\xi)$, leading to an empirical Bayes interpretation of fiducial distribution (4.11). We illustrate this idea in Example 4 in Section 6. It is also worth mentioning that though $\pi(\xi)$ is typically improper, the definition, of fiducial distribution as a conditional distribution of a random variable given another, guarantees that the the fiducial distribution in (4.11) is always proper.

Our interpretation of the fiducial recipe also sheds some new light on the consistency criterion of Lindley (1958). This criterion can be summarized as follows. Assume that the data is divided in two parts. If we calculate the fiducial distribution based on one part, use it as a prior and calculate the posterior based on the second part, we should get the same result as if we calculated the fiducial distribution based on all the data. Since fiducial distributions typically do not satisfy this criterion unless they are the same as a Bayesian posterior with respect to some prior, Lindley argued against the use of fiducial distributions. See Fraser (2006) for related discussion.

Notice that the proposed function $J(\mathbf{x}, \xi)$ depends on all the observations $\mathbf{x}$. Therefore the generalized fiducial distribution in (4.11) satisfies Lindley's consistency criterion if and only if $J(\mathbf{x}, \xi) = k(\mathbf{x})l(\xi)$, in which case the generalized fiducial distribution is the same as the Bayesian posterior with respect to the prior $l(\xi)$. On the other hand, if we used (4.10) as the generalized fiducial distribution, we would clearly satisfy Lindley's criterion as long as the order of the data remained the same. However, based on our experience from simulations, (4.11) usually has much better small sample frequentist properties than (4.10), and is therefore preferred even though it does not satisfy Lindley's consistency criterion.

Finally, the comparison between Bayesian and fiducial procedures for discrete data seems a little less clear to us. We comment on some aspects of this connection in Sections 6. and 7.

5. Asymptotic Consistency Results

In this section we present some general theorems that are applicable in situations one encounters in developing inference procedures using the generalized fiducial recipe. Consider a parametric statistical problem where we observe $X_1, \ldots, X_n$, whose joint distribution belongs to some family of distributions parametrized by $\xi \in \mathbb{R}^p$. We are interested in estimating $\theta = \pi(\xi) \in \mathbb{R}^q$. Let $\mathbb{S} = (S_1, \ldots, S_k)$, $k \geq q$, denote a statistic based on the X_i's. Denote by $\mathcal{R}_\theta(\mathbf{x}, U)$ a random variable having the distribution described in (4.3) and (4.4).

For simplicity of notation we define the following notion of convergence for open sets.

Definition 1. Sets A_n converge to an open set A, $A_n \to A$, if $(\lim A_n)^\circ = A$. Here $\lim A_n = B$ exists if $I_{A_n} \to I_B$, I_A is the indicator function of A, and B° is the interior of B.

We now state the conditions under which the generalized fiducial distribution defined in (4.3) leads to asymptotically correct frequentist coverage. We later show how these conditions can be verified in many applications.

Assumption 1.

1. There exist $\mathbf{t}(\xi) \in \mathbb{R}^k$ such that

$$\sqrt{n}\,(S_1 - t_1(\xi), \ldots, S_k - t_k(\xi)) \xrightarrow{\mathcal{D}} \mathbf{H} = (H_1, \ldots, H_k)^\top, \tag{5.1}$$

   where $\mathbf{H}$ has a non-degenerate multivariate normal distribution with mean 0 and variance Σ_H.

2. For each fixed $\mathbf{h} \in \mathbb{R}^k$, there is a random variable $R(\mathbf{h})$ such that
   (a) for any $\mathbf{x}_n \in \mathbb{R}^k$ satisfying $\sqrt{n}(\mathbf{x}_n - \mathbf{t}(\xi)) \to \mathbf{h}$, we have

$$\sqrt{n}(\mathcal{R}_\theta(\mathbf{x}_n) - \theta) \xrightarrow{\mathcal{D}} R(\mathbf{h});$$

   (b) there is a general matrix A and a non-negative definite matrix Σ_R such that $A\Sigma_H A^\top = \Sigma_R$, and $R(\mathbf{h})$ has multivariate normal distribution with mean $A\mathbf{h}$ and variance Σ_R.

3. $C(\mathbf{X}, \mathbf{z}, \mathbf{s}, \gamma) \subset \mathbb{R}^d$ is a collection of regions indexed by random variables $\mathbf{X}$, vectors $\mathbf{z} \in \mathbb{R}^d$, $\mathbf{s} \in \mathbb{R}^k$, and $\gamma \in (0,1)$ satisfying the following.
   (a) $C(\mathbf{X}, \mathbf{z}, \mathbf{s}, \gamma)$ is an open set with boundary of zero Lebesgue measure.
   (b) $P(\mathbf{X} \in C(\mathbf{X}, \mathbf{z}, \mathbf{s}, \gamma)) = \gamma$.
   (c) For all $a \in \mathbb{R}$ and $b \in \mathbb{R}^d$, $C(a\mathbf{X} + b, a\mathbf{z} + b, \mathbf{s}, \gamma) = aC(\mathbf{X}, \mathbf{z}, \mathbf{s}, \gamma) + b$.
   (d) For all $\mathbf{h} \in \mathbb{R}^k$, $\mathbb{Y}_n \xrightarrow{\mathcal{D}} R(\mathbf{h})$, $\mathbf{z}_n \to A\mathbf{h}$, $\mathbf{s}_n \to \mathbf{t}(\xi)$, and $\gamma_n \to \gamma$, we have $C(\mathbb{Y}_n, \mathbf{z}_n, \mathbf{s}_n, \gamma_n) \to C(R(\mathbf{h}), \mathbf{z}, \mathbf{t}(\xi), \gamma)$.

We now state our first theorem; its proof is relegated to the appendix.

Theorem 1. *Suppose Assumptions* 1 *holds and* $\gamma_n \to \gamma$. *Furthermore assume that there is a function* $\zeta : \mathbb{R}^k \to \mathbb{R}^d$ *such that for any* $\mathbf{s}_n \in \mathbb{R}^k$ *satisfying* $\sqrt{n}(\mathbf{s}_n - \mathbf{t}(\xi)) \to \mathbf{h}$, *we have*

$$\sqrt{n}(\zeta(\mathbf{s}_n) - \theta) \to A\mathbf{h}, \tag{5.2}$$

where the matrix A *is defined in* 2*b. Then* $\lim_{n\to\infty} P_\xi\left(\theta \in C(\mathcal{R}_\theta(\mathbb{S}), \zeta(\mathbb{S}), \mathbb{S}, \gamma_n)\right) = \gamma$. *In particular* $C(\mathcal{R}_\theta(\mathbb{S}), \zeta(\mathbb{S}), \mathbb{S}, \gamma)$ *is a confidence region for* θ *with asymptotic coverage probability equal to* γ.

Remark 3. This theorem is in truth a theorem about the choice of the region $C(\mathbf{X}, \gamma)$. There are many regions of probability $1-\gamma$ available; Condition 3 gives shapes of good regions $C(X, \gamma)$.

For example, if $d = 1$, one of the typical choices is the upper confidence region, $C(X, \gamma) = (-\infty, q(X, \gamma))$, where $q(\mathbf{X}, \gamma)$ is the γ-quantile of the distribution of X. Other choices are the lower confidence region, and the two sided, equal tailed region. If $d > 1$, we can also consider the equal tailed regions. In fact the conditions on the region are so flexible that they allow for most typical multiple comparison regions. We demonstrate this in Example 9 in Section 6.

The more important part of the story is hidden in Assumptions 1, 2. These conditions ensure that the generalized fiducial distribution satisfies the Bernstein-von Mises theorem—see Chapter 8 of Le Cam and Yang (2000). This happens quite often. The exact assumptions under which the Bernstein-von Mises theorem holds for generalized fiducial quantities are the subject of ongoing investigation. In what follows we present Theorem 2 as a first step in this direction.

Remark 4. It is fairly straightforward to generalize the statements of Theorem 1 for distributions that are not in the domain of attraction of the normal distribution. Some examples in that direction have been explored in

Hannig, Iyer and Patterson (2006b). However, the main ideas are better demonstrated within the setting we have chosen. In particular the key Condition 2b is easier to understand if the limiting distribution is normal.

To conclude this section we show that the Assumptions 1 and 2 are satisfied in the basic regular case. Assume that $X_1, X_2, \dots$ are i.i.d. with density $f_\theta = f(x,\theta)$ and distribution function $F(x,\theta)$, where $\theta \in \Theta$ and Θ is an open subset of $\mathbb{R}$.

We assume the usual regularity assumptions on f_θ. They can be found for example in Ghosh and Ramamoorthi (2003, pp.34-35), assumptions (i)-(v). Since they are standard, we do not list them. We also assume an additional regularity condition.

(v') Set $L_n(\theta) = \sum_{i=1}^n \log f(X_i,\theta)$ and assume that for any $\delta > 0$,

$$\inf_{|\theta-\theta_0|>\delta} \frac{\min\limits_{i=1,\dots,n} \log f(X_i,\theta)}{|L_n(\theta) - L_n(\theta_0)|} \xrightarrow{P_{\theta_0}} 0.$$

Notice that typically, for a fixed θ_0, the numerator goes to negative infinity as $-C_1 \log n$ while the denominator goes to negative infinity like $-C_2 n$. If this happens, the assumption (v') guarantees that C_1/C_2 remains bounded as a function of θ.

It is well-known, Theorem 1.4.1 of Ghosh et al. (2003), that there is a statistic (usually called the maximum likelihood estimator) S_n such that $\sqrt{n}(S_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, 1/I(\theta_0))$, where $I(\theta_0)$ is the Fisher information. Condition 1 is thereby satisfied.

We now show that the Condition 2 is satisfied as well. Choose the structural equation $X_i = F^{-1}(\theta, U)$ and assume that the various invertibility conditions of Section 4.1 are satisfied. The fiducial distribution is then given by (4.13) and (4.14). For future reference we denote the density of $\sqrt{n}(\mathcal{R}_\theta - S)$ by $\pi_r(\theta,\mathbf{x})$. Our second theorem is actually stronger than what is required to verify Condition 2; the proof is relegated to the appendix.

Theorem 2. *Adopt the regularity conditions* (i)$-$(v) *of* Ghosh et al. (2003, pp.34-35), (v') *of above and the invertibility conditions of Section* 4.1. *Assume further that* $J(x,\bullet)$ *is continuous in* θ, $\pi(\theta) = E_{\theta_0} J_0(X,\theta)$ *is finite,* $\pi(\theta_0) > 0$ *and, on some neighborhood of* θ_0,

$$E_{\theta_0}\left(\sup_{\theta\in(\theta_0-\delta_0,\theta_0+\delta_0)} J_0(X,\theta)\right) < \infty.$$

Then

$$\int_{\mathbb{R}} \left| \pi_r(\theta,\mathbf{x}) - \frac{e^{-s^2/(2/I(\theta_0))}}{\sqrt{2\pi/I(\theta_0)}} \right| \xrightarrow{P_{\theta_0}} 0. \tag{5.3}$$

Remark 5. The main idea of the proof is to use the Uniform Strong Law of Large Numbers to show that the quantity $J(\mathbf{x}, \theta)$ converges to $\pi(\theta)$ uniformly on compacts in θ, and then use this to prove that the difference between the generalized fiducial distribution and the posterior with respect to the prior $\pi(\theta)$ is negligible. The statement then follows by the regular Bernstein-von Mises theorem for Bayesian posteriors.

A similar idea will apply even if the parameter space is p-dimensional, in which case the quantity $J(\mathbf{x}, \xi)$ is a U-statistics. One can then replace the Uniform Strong Law by the Uniform Strong Law of Large Numbers for U-statistics Yeo and Johnson (2001). The rest of the proof is basically the same.

6. Examples

The purpose of this section is to explain the use of the generalized fiducial recipe on several examples of varying complexity.

Variance Component Model

Example 4. The first example is motivated by an unbalanced variance components model. Such models arise in heritability studies in animal breeding experiments Burch and Iyer (1997), quality improvement studies in manufacturing processes Burdick, Borror and Montgomery (2005a), characterizing sources of error in general variance components models Liao and Iyer (2004), and in many other applications. In the simplest case one has the model

$$Y_{ij} = \mu + A_i + e_{ij},$$

where μ is an unknown parameter, the A_i are i.i.d. $N(0, \phi)$, the e_{ij} are i.i.d. $N(0, \theta)$, and all random variables are jointly independent. In metrology, Y_{ij} might be the diameter measurement on a part (ball-bearing) and μ the mean diameter of the population of ball-bearings output by the process. A random sample of a ball-bearings is selected. The true diameter of the i^{th} ball-bearing is $\mu + A_i$. Ball-bearing i is measured n_i times. If $n_i = n$ for all i we have a balanced one-way random effects model; in the case of unequal n_i, we have an unbalanced one-way random model. In the balanced case the complete sufficient statistics are well-known Searle, Casella and McCulloch (1992). In the unbalanced case the minimal sufficient statistics are incomplete. Inference about ϕ and θ is typically based on K independent quadratic forms that have scaled chi-square distributions and whose expected values have the form $\theta + c_i\phi$ for some known c_i, $i = 1, \ldots, K$. Hence we illustrate our procedure by obtaining a generalized fiducial distribution for (ϕ, θ).

We use the structural equation

$$S_i = \frac{(c_i\phi + \theta)U_i}{n_i}, \quad i = 1, \dots, K,$$

where $c_1 > \cdots > c_K \geq 0$, and $U_1, \dots, U_K$ are independent chi-square random variable with $n_1, \dots, n_K$ numbers of degrees of freedom, respectively.

Since this structural equation satisfies the various assumptions in Section 4.1, we can use the formula in (4.11). Recall that the recipe suggest taking two equations to solve for ϕ and θ and use the rest for conditioning. In particular if equations i, j are chosen,

$$\mathbf{G}^{-1}_{0,(i,j)}(\mathbf{S}, (\phi, \theta)) = \begin{pmatrix} \frac{n_i s_i}{c_i\phi+\theta} \\ \frac{n_j s_j}{c_j\phi+\theta} \end{pmatrix}$$

leading to

$$\begin{aligned} J_{0,(i,j)}(\mathbf{s}, (\phi, \theta)) &= \left| \frac{\det \begin{pmatrix} -c_i n_i s_i/(c_i\phi+\theta)^2 & -n_i s_i/(c_i\phi+\theta)^2 \\ -c_i n_j s_j/(c_j\phi+\theta)^2 & -n_j s_j/(c_j\phi+\theta)^2 \end{pmatrix}}{\det \begin{pmatrix} n_i/(c_i\phi+\theta) & 0 \\ 0 & n_j/(c_j\phi+\theta) \end{pmatrix}} \right| \\ &= \frac{|c_i - c_j| s_i s_j}{(c_i\phi+\theta)(c_j\phi+\theta)}. \end{aligned}$$

The fiducial distribution (4.11) then becomes

$$f_{\mathcal{R}_{\phi,\theta}}(\phi, \theta) = C^{-1} \frac{\exp\Big(-(1/2)\sum_{i=1}^{d} n_i s_i/(c_i\phi+\theta) \Big)}{\prod_{i=1}^{d}(c_i\phi+\theta)^{n_i/2}} J(\mathbf{s}, (\phi, \theta)), \tag{6.1}$$

where C is a normalizing constant and

$$J(\mathbf{s}, (\phi, \theta)) = \binom{K}{2}^{-1} \sum_{i \neq j} \frac{|c_i - c_j| s_i s_j}{(c_i\phi+\theta)(c_j\phi+\theta)}.$$

To set up confidence regions one can use numerical integration. The main parameter of interest in this situation is ϕ. The fiducial distribution of ϕ does not lead to exact frequentist inference. However, simulation results suggest good properties (as measured by coverage and length of confidence intervals based on the fiducial distribution). Moreover, it can be shown that the fiducial distribution leads to asymptotically correct frequentist inference. If K is fixed, one can prove it directly using Theorem 1. However, the asymptotics needed in most practical applications involve letting $K \to \infty$. The proof in this more complicated setting goes well beyond the scope of this paper and the reader is referred

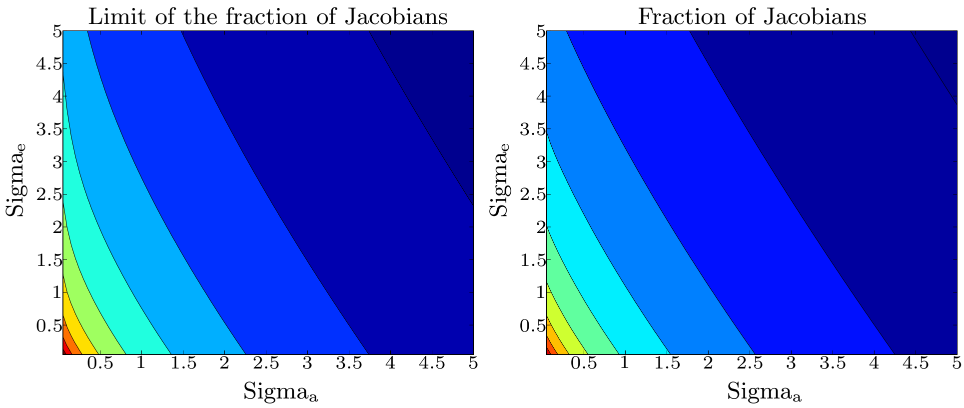

Figure 1. Contour plots of $J(\mathbf{s},(\phi,\theta))$ and $\pi(\phi,\theta)$ when $K=4$, $\phi_0=1$, $\theta_0=0.2$, $c_1=0.01, n_1=1$, $c_2=0.5, n_2=2$, $c_3=1.1, n_3=4$ and $c_4=4, n_4=1$. Shows that $J(\mathbf{s},(\phi,\theta)$ is a reasonable estimator of $\pi(\phi,\theta)$ even for small n_i.

to E, Hannig and Iyer (2008), where an interested reader can find the proof, simulation study results, and additional discussion.

We end this example with an investigation of the quantity $J(\mathbf{s},(\phi,\theta))$ that plays a role of a "data dependent prior" in (6.1). Assume that K is fixed, (ϕ_0,θ_0) are the "true parameters" used in generating $\mathbf{s}$, and for all $i=1,\ldots,K$, $n_i\to\infty$. Strong Law of Large Numbers implies that $s_i\to(c_i\phi_0+\theta_0)$ *a.s.*, and

$$J(\mathbf{s},(\phi,\theta))\to\pi(\phi,\theta)=\binom{K}{2}^{-1}\sum_{i\neq j}\frac{|c_i-c_j|(c_i\phi_0+\theta_0)(c_j\phi_0+\theta_0)}{(c_i\phi+\theta)(c_j\phi+\theta)}\ a.s..$$

The limit $\pi(\phi,\theta)$ is a function of (ϕ,θ) and does not depend on the data. It can therefore be considered as a prior. However, this is not feasible as it depends also on the unknown true parameters. As discussed above $J(\mathbf{s},(\phi,\theta))$ could be considered as a data-based estimator of the infeasible prior $\pi(\phi,\theta)$. We illustrate this in Figure 1, where we show a contour plot of one realization of the random quantity $J(\mathbf{s},(\phi,\theta))$ and a contour plot of the deterministic $\pi(\phi,\theta)$ when $K=4$; $\phi_0=1$, $\theta_0=0.2$; $c_1=0.01, n_1=1$; $c_2=0.5, n_2=2$; $c_3=1.1, n_3=4$ and $c_4=4, n_4=1$. The plot demonstrates that the two functions show a surprising level of agreement even though the values of the n_i are very small.

Fiducial inference for the multinomial distribution

The next series of examples considers generalized fiducial inference for the multinomial distribution on $k+1$ categories $\{1,\ldots,k+1\}$. The special case of the binomial distribution ($k=1$) has received some recent attention by Brown, Cai and DasGupta (2001, 2002) and Cai (2005). These authors show that the classical solutions based on normal approximations do not have good small sample properties and they recommend some alternative solutions. The one recommendation that stands out consistently is the interval estimate based on the posterior distribution arising from the Jeffreys prior. Later we show that there is, in fact, one of the fiducial intervals. We also show that there is another fiducial solution for the binomial parameter p that does just as well.

Example 5. Let $X_1, \ldots, X_n$ be i.i.d. multinomial($\boldsymbol{p}$) random variables, where $\boldsymbol{p} = (p_1, \ldots, p_k)$, $p_j \in [0,1], j = 1, \ldots, k$, and $\sum_{j=1}^k p_j \leq 1$. We derive a generalized fiducial distribution for $\boldsymbol{p}$. Set $q_0 = 0$ and $q_j = \sum_{l=1}^j p_l$, $j = 1, \ldots, k$. The structural equations for the $X_i, i = 1, \ldots, n$, can be expressed as

$$X_i = \sum_{j=0}^{k} I_{[q_j,1]}(U_i), \tag{6.2}$$

where $U_1, \ldots, U_n$ are i.i.d. U$(0,1)$ random variables.

Assume that we have observed $x_1, \ldots, x_n$, and denote the number of occurrences of j by n_j. For $j = 1, \ldots, k+1$, define $t_j = \sum_{r=1}^j n_r$. In particular, $t_{k+1} = n$. Let $U_{s:n}$ denote the s^{th} order statistic among $U_1, \ldots, U_n$. For simplicity of notation define $t_0 = 0$, $U_{0:n} = 0$ and $U_{n+1:n} = 1$. The set $Q(\mathbf{x}, \mathbf{U}) \neq \emptyset$ if and only if

$$n = \sum_{j=1}^{k+1} \sum_{i=1}^{n} I(X_i = j) I\left(U_i \in (U_{t_{j-1}:n}, U_{t_j:n}]\right).$$

In this case $Q(\mathbf{x}, \mathbf{U}) = Q^\star(\mathbf{x}, \mathbf{U})$, where

$$Q^\star(\mathbf{x}, \mathbf{U}) = \left\{ (p_1, \ldots, p_k) \middle| (q_1, \ldots, q_k) \in \bigtimes_{j=1}^{k} \left(U_{t_j:n},\, U_{t_j+1:n}\right] \right\}.$$

Here $\bigtimes_i A_i$ is the cartesian product of the sets A_i and q_i is as in (6.2). In particular for $j = 1, \ldots, k$, $p_j = q_j - q_{j-1}$ and $p_{k+1} = 1 - q_k$.

The exchangeability of $U_i, i = 1, \ldots, n$, then implies that the conditional distribution of $V(Q(\mathbf{x}, \mathbf{U}))$, conditional on the event $Q(\mathbf{x}, \mathbf{U}) \neq \emptyset$, is the same as the (unconditional) distribution of $V(Q^\star(\mathbf{x}, \mathbf{U}))$. By our definition the generalized fiducial quantity is $\mathcal{R}_{\boldsymbol{p}}(\mathbf{x}) = V(Q^\star(\mathbf{x}, \mathbf{U}))$. Equivalently there is a random vector $\mathbf{D} = (D_1, \ldots, D_k)$ with support $[0,1]^k$ such that

$$\mathcal{R}_{\boldsymbol{p}}(\mathbf{x}) = (R_1, R_2 - R_1, \ldots, R_k - R_{k-1})^\top, \tag{6.3}$$

where $R_j = U_{t_j:n} + D_j(U_{t_j+1:n} - U_{t_j:n})$.

Notice that if $n_j = 0$ for some $j = 2, \ldots, k$, it would be possible to get a negative value for $\mathcal{R}_{p_i}$ the i^{th} element of $\mathcal{R}_{\mathbf{p}}$. This can be prevented by requiring the random vector $\mathbf{D}$ to satisfy $D_j \geq D_{j-1}$ whenever $n_j = 0$.

The observation made in the previous paragraph implies that the generalized fiducial distribution depends on the particular choice of the structural equation (6.2). In particular, if one or more categories are not observed in our sample, we might get a different generalized fiducial distribution by relabeling.

Using Theorem 1, we now show that inference based on $\mathcal{R}_{\boldsymbol{p}}(s)$ has good frequentist properties asymptotically. Since we consider equal-tailed regions based on the distribution of $\mathcal{R}_{\boldsymbol{p}}(s)$, Assumption 1.3 is automatically satisfied. Let S_j be the number of times we observe value j among the $X_1, \ldots, X_n$. Recall that $\mathbb{S} = (S_1, \ldots, S_k)^\top$ has a multinomial$(n, p_1, \ldots, p_k)$ distribution. Therefore $\sqrt{n}(\mathbf{S}/n - \boldsymbol{p}) \to H$, where $H \sim \mathrm{N}(0, \Sigma)$ and $\Sigma = \mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top$. This verifies Assumption 1.1.

Notice that for any sequence of integers k_j, where $0 \leq k_j \leq j$, we have $n(U_{k_n+1:n} - U_{k_n:n}) \xrightarrow{\mathcal{D}} \Gamma(1,1)$. Fix $\mathbf{h}$, set $\mathbf{s}/n = (\boldsymbol{p} + \mathbf{h}/\sqrt{n})$, and write $\mathbf{W}_n = (U_{s_1:n}, U_{s_1+s_2:n}, \ldots, U_{s_1+\cdots+s_k:n})$. A simple calculation shows that $\sqrt{n}(\mathbf{W}_n - \mathbf{q}) \xrightarrow{\mathcal{D}} N(\mathbf{g}, \widehat{\Sigma})$, where $g_j = \sum_{k=1}^{j} h_k$ and $\widehat{\Sigma}_{i,j} = \min(q_i, q_j)(1 - \max(q_i, q_j))$, with $q_j = \sum_{k=1}^{j} p_k$. Thus, by Slutsky's theorem, $\sqrt{n}(\mathcal{R}_{\boldsymbol{p}}(S) - \boldsymbol{p}) \xrightarrow{\mathcal{D}} N(\mathbf{h}, \Sigma)$. All of Assumption 1 is satisfied. In particular we can conclude that generalized fiducial confidence sets will have asymptotically correct frequentist coverage regardless of the choice of the distribution $V(\,\boldsymbol{.}\,)$.

We further investigate this generalized fiducial quantity in two special cases, the binomial distribution ($k = 1$) and the trinomial distribution ($k = 2$).

Special case 1 - The binomial distribution.

Example 6. For the special case of a binomial distribution, a generalized fiducial quantity for p is

$$\mathcal{R}_p(x) = U_{s:n} + D(U_{s+1:n} - U_{s:n}), \tag{6.4}$$

with D being any random variable with support contained in $[0, 1]$, and s being the observed number of successes.

Recall that the joint density of $(U_{s:n}, U_{s+1:n})$ is

$$f_{(U_{s:n}, U_{s+1:n})}(u, v) = \frac{n!}{(s-1)!(n-s-1)!} u^{s-1}(1-v)^{n-s-1}, \; 0 < u < v < 1.$$

Therefore, the density of $\mathcal{R}_p$ is

$$f_{\mathcal{R}_p}(p) = \int_0^1 \int_0^{\frac{p}{d} \wedge \frac{1-p}{1-d}} \binom{n}{s} s(p - dq)^{s-1}$$

$$\times (n-s)\left((1-p)-(1-d)q\right)^{n-s-1}\, dq\, dF_D(d)\, I_{(0,1)}(p), \quad (6.5)$$

where $F_D(d)$ is the distribution function of D and $x \wedge y = \min\{x, y\}$. If, additionally, D is continuous with density f_D, (6.5) simplifies to

$$f_{\mathcal{R}_p}(p) = \binom{n}{s} \int_0^p \int_p^1 f_D\left(\frac{p-u}{v-u}\right) \frac{s u^{s-1}(n-s)(1-v)^{n-s-1}}{v-u}\, dv\, du\, I_{(0,1)}(p). \quad (6.6)$$

There are many reasonable choices for the distribution of D in the description of $\mathcal{R}_p$. We have considered five different choices that appeared natural to us. For the first three choices we took D to be random and independent of $U_1, \ldots, U_n$.

First, the maximum entropy choice is $D \sim$ uniform$(0, 1)$.

Second, the maximum variance choice, suggested implicitly by Efron (1998), is $D \sim$ uniform$\{0, 1\}$. We remark that a direct calculation, cf. Grundy (1956), shows that these two choices lead to a generalized fiducial distribution that is not a Bayesian posterior with respect to any prior.

The third choice $D \sim$ Beta$(1/2, 1/2)$ leads to $\mathcal{R}_p \sim$ Beta$(s+1/2, n-s+1/2)$, which is the Bayesian posterior for Jeffreys prior.

The fourth choice is a little harder to describe in terms of D. It is $\mathcal{R}_p \sim$ Beta$(s+1, n-s+1)$. This is the scaled likelihood, or posterior with respect to the flat prior. Beta$(s+1, n-s+1)$ is a generalized fiducial distribution according to our definition, since it is stochastically larger than the distribution of $U_{s:n}$, which is Beta$(s+1, n-s)$, and stochastically smaller than the distribution of $U_{s+1:n}$, which is Beta$(s, n-s+1)$. This can be seen by noticing that, conditional on $U_1, \ldots, U_n$, the distribution of D is given by $D = 0$ with probability $U_{s:n}$, $D = 1$ with probability $1 - U_{s+1,n}$, and $D \sim U(0, 1)$ with probability $U_{s+1:n} - U_{s:n}$.

The last choice $D = 1/2$, corresponds to the midpoint of the interval $(U_{s:n}, U_{s+1:n})$.

To evaluate the performance of the generalized fiducial distribution and compare the performance of the various choices of D, we carried out an extensive simulation study. As we have seen earlier, generalized fiducial inference for the multinomial distribution is asymptotically correct. Therefore, our simulation study concentrated mostly on small values of n. In particular, we considered $n = 3, 6, 9, \ldots, 45, 48, 100, 1,000$ and $p = 0.01, 0.02, \ldots, 0.99$. For each of the combinations of n and p we simulated $5,000$ evaluations of the probability $Q(\mathbb{X}) = P(\mathcal{R}_p(\mathbb{X}) < p|\mathbb{X})$ using each of the five variations of generalized fiducial distribution. If the generalized fiducial inference were exact, the $Q(\mathbb{X})$ should follow the U$(0, 1)$ distribution. The level of agreement of $Q(\mathbb{X})$ with the U$(0, 1)$ distribution was examined using QQ-plots.

Since generalized fiducial inference is a non-randomized procedure, the distribution of $Q(\mathbb{X})$ can take only n values. Therefore it cannot be expected that the agreement with the uniform distribution would be very good for small values

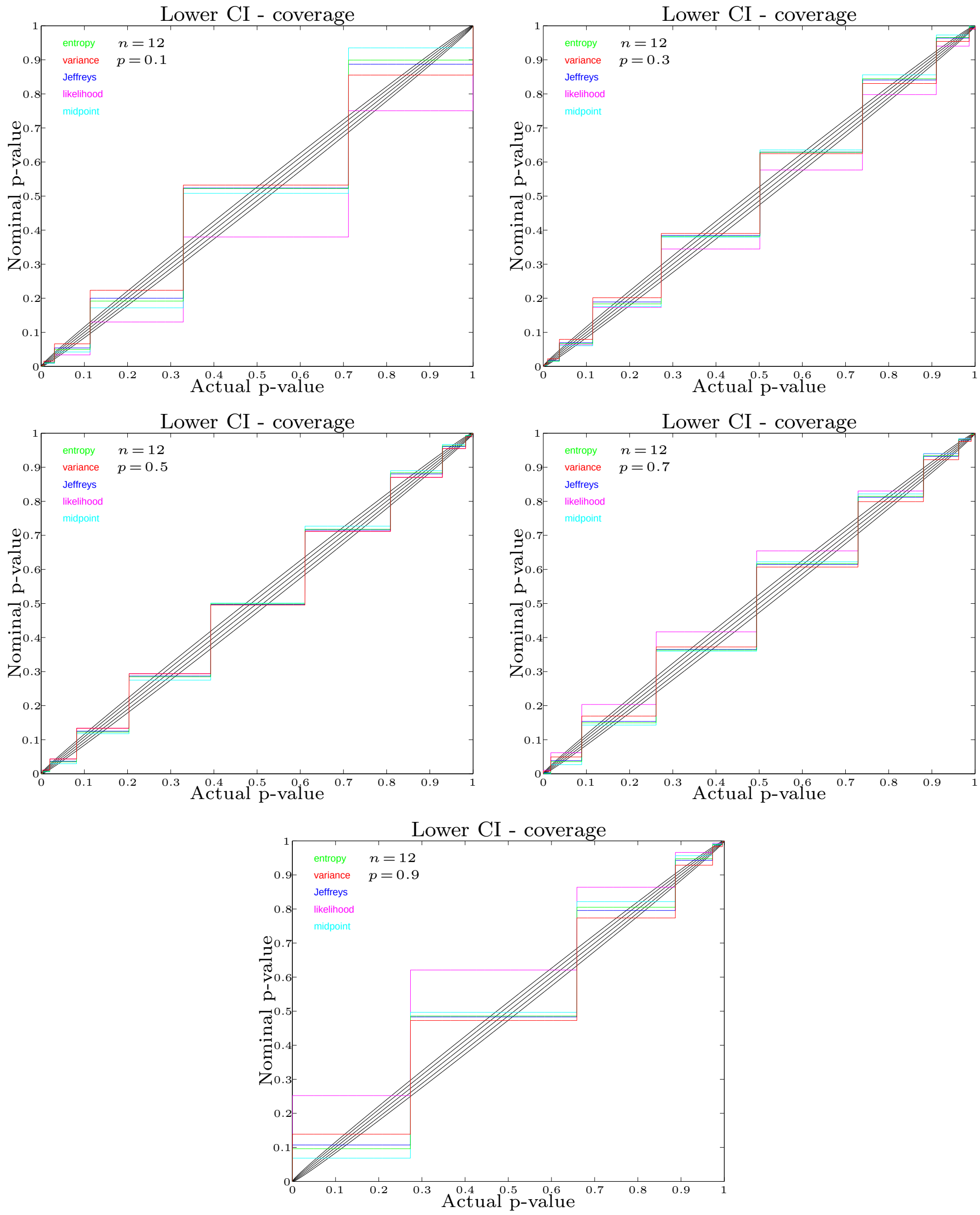

Figure 2. QQ-plots of $Q(\mathbb{X})$ for $n = 12$ and $p = 0.1, 0.3, 0.5, 0.7, 0.9$. The black color correspond to an area of natural fluctuation of a QQ-plot due to randomness; the colored graphs correspond to the QQ-plots of the various generalized fiducial distributions.

of n. However, the agreement improves dramatically as n increases. To illustrate this, we show the QQ-plots for $n = 12$ and $p = 0.1, 0.3, 0.5, 0.7, 0.9$ in Figure 2;

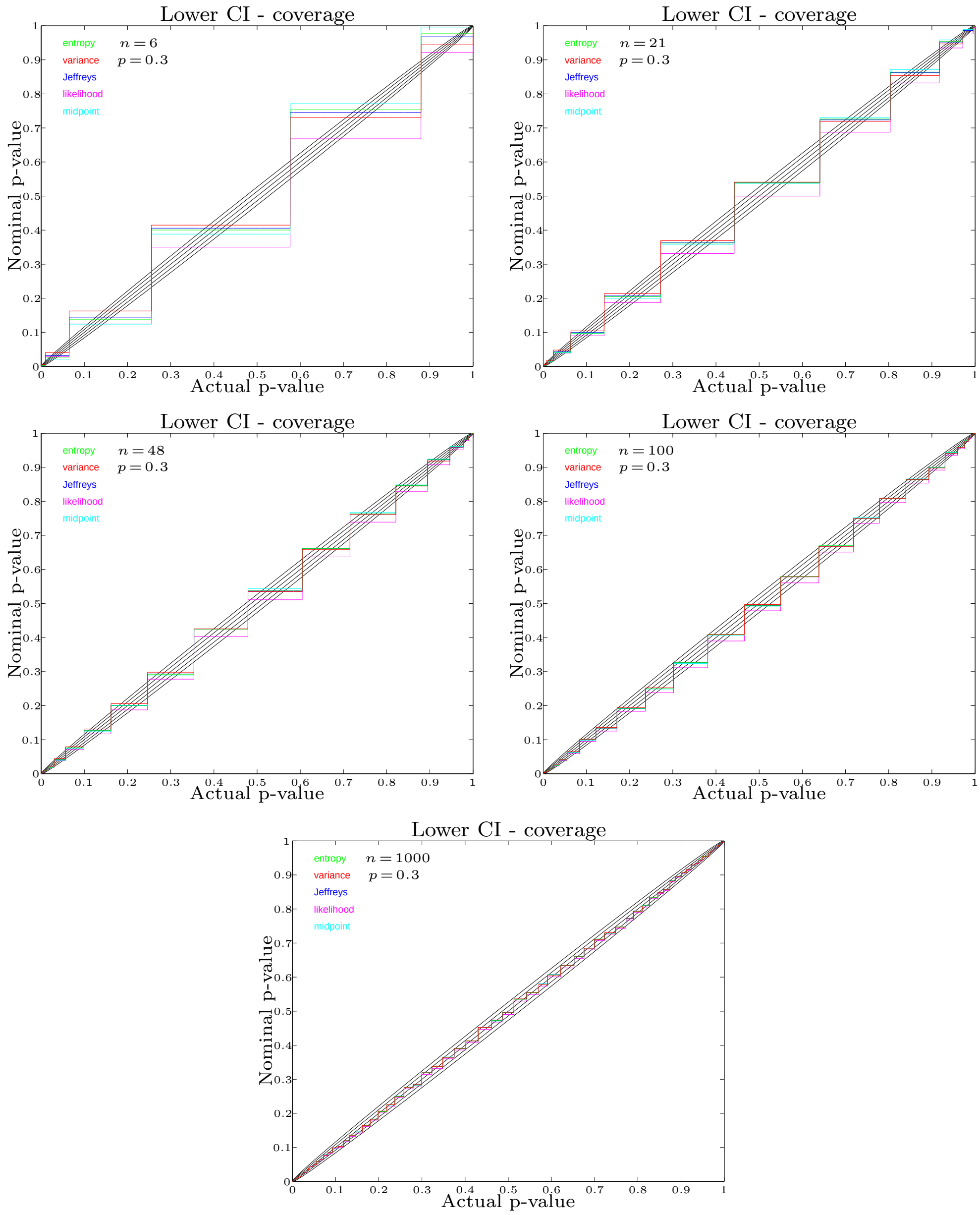

Figure 3. QQ-plots of $Q(\mathbb{X})$ for $n = 6, 21, 48, 100, 1,000$ and $p = 0.3$. The black color correspond to an area of natural fluctuation of a QQ-plot due to randomness; the colored graphs correspond to the QQ-plots of the various generalized fiducial distributions.

we show QQ-plots for $n = 6, 21, 48, 100, 1,000$ and $p = 0.3$ in Figure 3.

The closer the points on the QQ-plot are to the line $y = x$, the better the performance of the procedure. We can see straightaway that the scaled likelihood performs worse than any of the other choices. To make this comparison more rigorous we compute, for each of the choices of D,

$$A = \int_0^1 |F_Q(x) - x|\, dx, \quad \text{and} \quad D = \int_0^1 (x - F_Q(x))\, dx,$$

where $F_Q(x)$ is the empirical distribution function of the observed values of the $Q(\mathbb{X})$. Smaller values of A and D signify better overall fit. Since we are planning to use the generalized fiducial distribution for inference, one can argue that the center of the distribution of $Q(\mathbb{X})$ is of little importance. Therefore we also check for the level of agreement in the tails. To this end, let

$$A_l = \int_0^{.1} |F_Q(x) - x|\, dx, \quad D_l = \int_0^{.1} (x - F_Q(x))\, dx,$$
$$A_u = \int_{.9}^1 |F_Q(x) - x|\, dx, \quad \text{and} \quad D_u = \int_{.9}^1 (F_Q(x) - x)\, dx.$$

Here we chose A_l, D_l to describe the average fit for typical lower tail CIs, and A_u, D_u to describe the average fit for typical upper tail CIs. In both cases positive values of D_l and D_u correspond to being conservative, while negative values of D_l and D_u correspond to being anticonservative.

For each fixed n we plotted the graphs of these statistics as functions of the probability p. For illustration we show plots of of these quantities for $n = 6, 21, 48, 50, 100$ in Figures 4, 5 and 6.

One finds that the best choice is the maximum variance choice of $D \sim$ uniform$\{0,1\}$ is consistently better than other choices. However, $D \sim U(0,1)$ and $D \sim \mathrm{B}(1/2, 1/2)$ (the maximum entropy and posterior with respect to Jeffreys prior) were typically very close to it. The last two choices were found to perform not as well. In particular, the scaled likelihood underperformed the other choices by a large margin. In light of this we recommend the choice $D \sim$ uniform$\{0,1\}$.

Remark 6. Cai (2005) has investigated the two-term Edgeworth expansions for coverage of several one-sided Binomial Confidence Intervals. We remark that similar calculations can be used to derive the two-term Edgeworth expansion for the generalized fiducial distributions discussed here. In particular one can show that, just like confidence intervals based on posteriors calculated using Jeffreys prior, the maximum variance generalized fiducial distribution leads to confidence intervals that are first order matching, see Ghosh (1994).

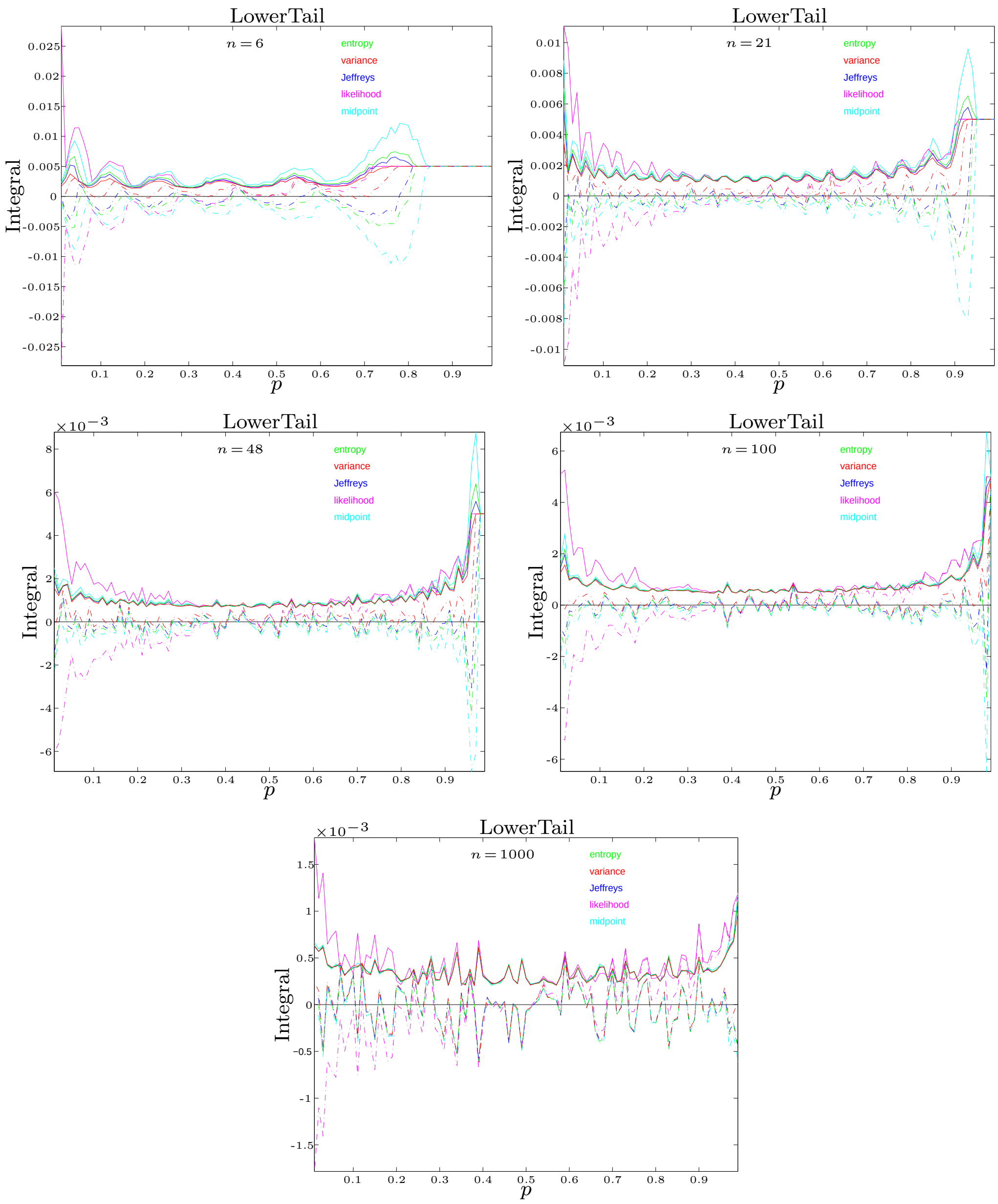

Figure 4. Plots of A_l (solid line) and D_l (dashed line) as functions of p for $n = 6, 21, 48, 100, 1,000$. Small values of A_l and D_l are preferable. Positive values of D_l correspond to the method being conservative on average. The various colors correspond to various choices for the generalized fiducial distribution.

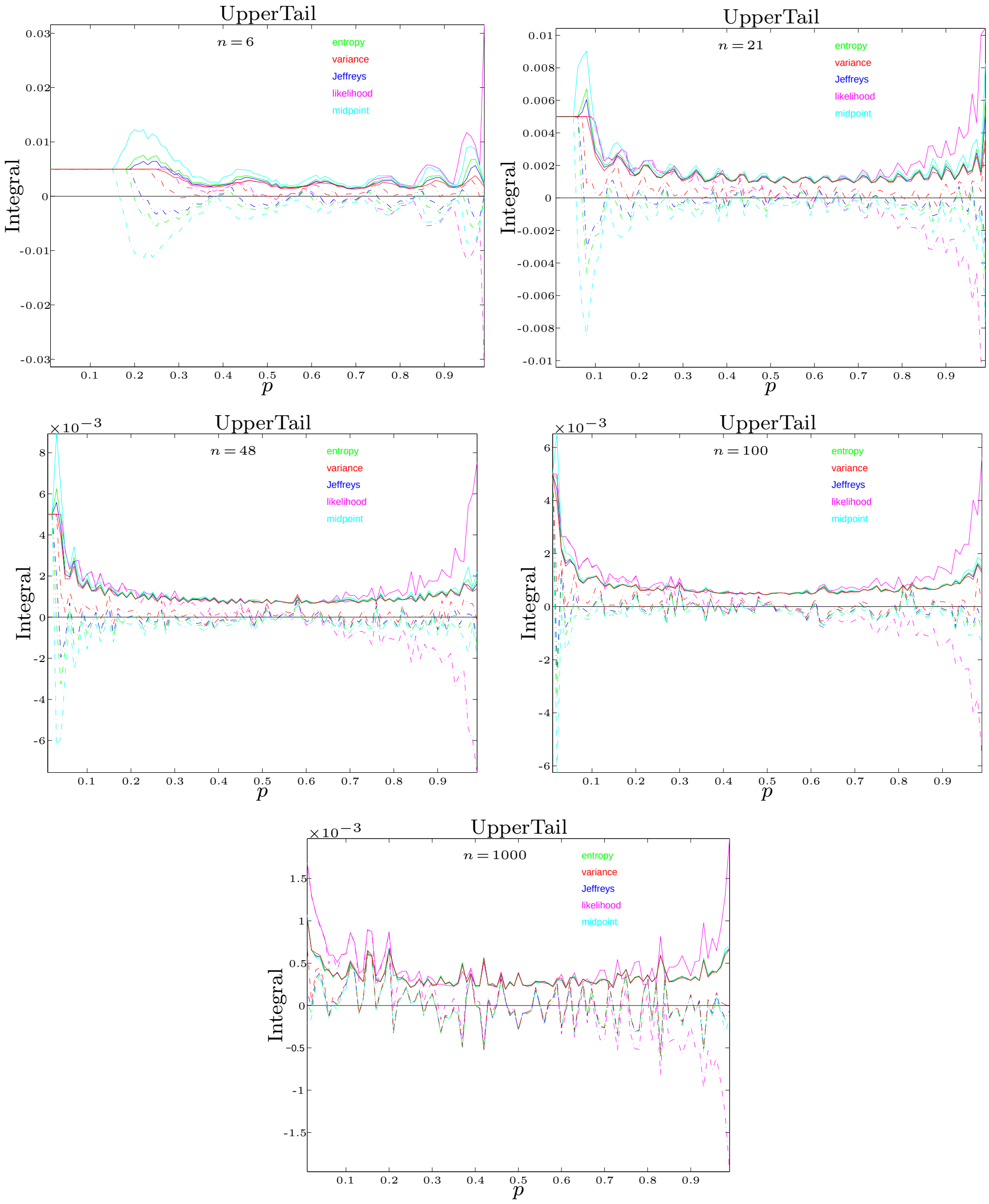

Figure 5. Plots of A_u (solid line) and D_u (dashed line) as functions of p for $n = 6, 21, 48, 100, 1,000$. Small values of A_u and D_u are preferable. Positive values of D_u correspond to the method being conservative on average. The various colors correspond to various choices for the generalized fiducial distribution.

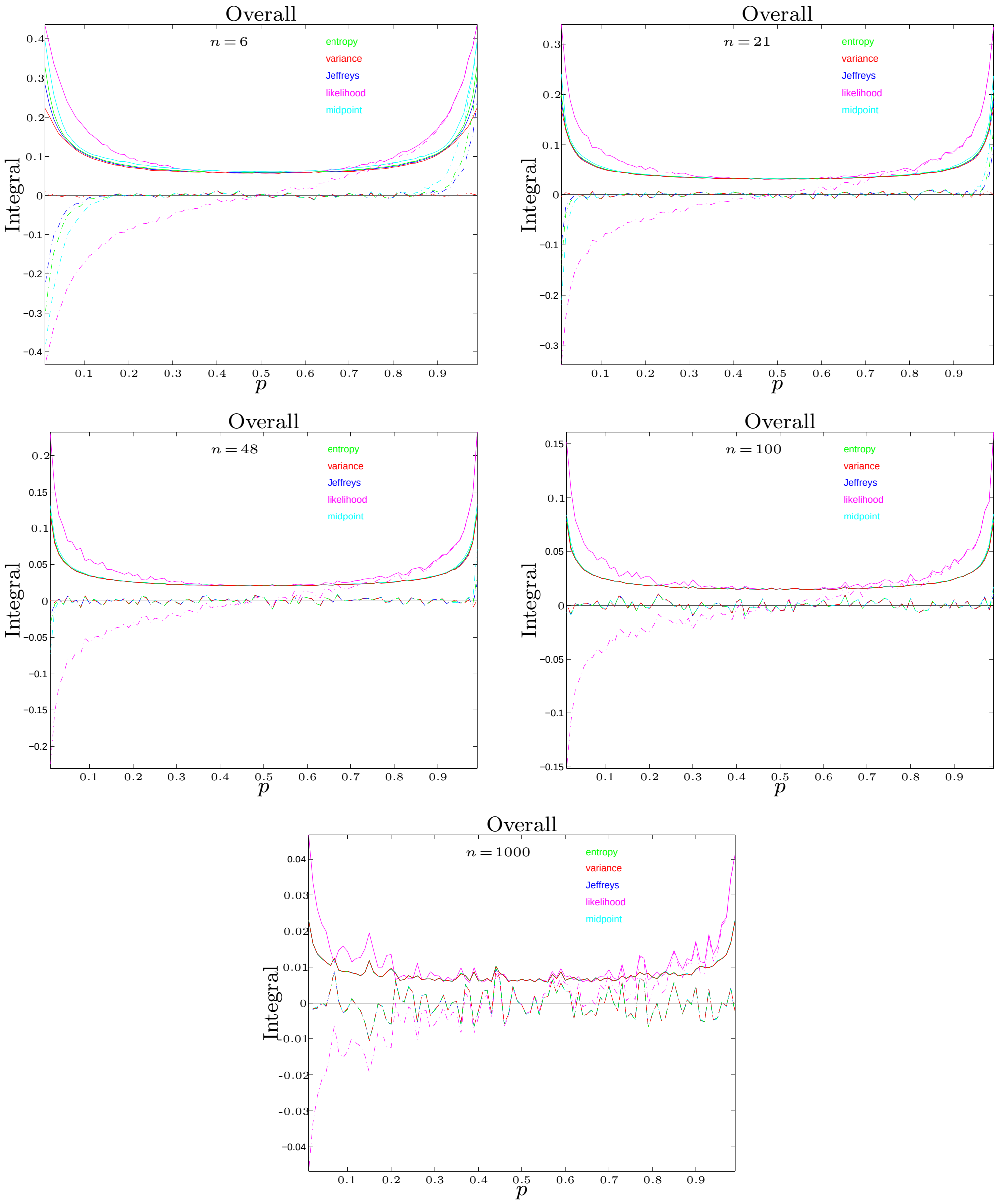

Figure 6. Plots of A (solid line) and D (dashed line) as functions of p for $n = 6, 21, 48, 100, 1,000$. Small values of A and D are preferable. The various colors correspond to various choices for the generalized fiducial distribution.

Special case 2 - The trinomial distribution.

Example 7. Some aspects of the fiducial distribution for the parameters of a trinomial have been investigated by Dempster (1968); he used a trinomial distribution as an example for his definition of upper and lower probabilities. In this example we investigate the small sample frequentist properties of the generalized fiducial distribution for the trinomial parameters. There are many reasonable choices for the distribution of $\mathbf{D}$ in (6.3). We considered five choices that appeared natural. Based on our experience from Example 6, we take $\mathbf{D}$ independent of $U_1, \ldots, U_n$.

The maximum entropy choice is achieved by taking $\mathbf{D}$ to be $\mathrm{U}(0,1)^2$ if $s_2 > 0$, and $\mathbf{D} \sim \mathrm{U}\{(x,y), 0 < x < y < 1\}$ if $s_2 = 0$.

The Bayesian posterior for the Jeffreys prior is achieved by taking D_1, D_2 as i.i.d. Beta$(1/2, 1/2)$ if $s_2 > 0$, $D_1 \sim$ Beta$(1/2, 1/2)$, and $D_2 = 1$ if $s_2 = 0$.

The third choice is a version of a maximum variance distribution. Here $\mathbf{D} \sim$ uniform$\{0,1\}^2$ if $s_2 > 0$, and $\mathbf{D} \sim$ uniform$\{(0,0),(0,1),(1,1)\}$ if $s_2 = 0$. This is obtained by maximizing the determinant of the covariance matrix of $\mathcal{R}_{\mathbf{p}}(\mathbf{x})$. Notice that it is also the uniform distribution on the vertices of $Q(\mathbf{x}, U)$.

The fourth choice is another version of a maximum variance distribution, it is obtained by maximizing the smallest eigenvalue of the covariance matrix of $\mathcal{R}_{\mathbf{p}}(\mathbf{x})$. Notice that this distribution is supported on the vertices of $Q(\mathbf{x}, U)$.

Our last choice is the uniform distribution on the boundary of $Q(\mathbf{x}, U)$.

To evaluate the performance of the generalized fiducial distribution and to compare the performance of the various choices of $\mathbf{D}$, we performed an extensive simulation study concentrated on small values of n. In particular we considered $n = 5, 10, 15, \ldots, 30, 300$ and $p_1, p_2 \in \{0.05, 0.1, \ldots, 0.95\}$ with $p_1 + p_2 < 1$. For each of the combination of the parameters n, p_1, p_2, we simulated a sample of $2,000$ observations from the trinomial distribution. For each of the trinomial observation and each of the choice of $\mathbf{D}$ we generated a sample of $3,000$ observations from the generalized fiducial distribution $\mathcal{R}_{\mathbf{p}}(\mathbf{x})$.

In order to evaluate the quality of the joint generalized fiducial distribution we then evaluated the empirical coverage of the one-sided equal tailed region. In particular, for any random vector $\mathbb{X}$ and $0 < \alpha < 1$, we define the one-sided equal-tailed region $C(\mathbb{X}, \alpha)$ as the set $\{(x.y); x \le x_0, y \le y_0\}$ satisfying $P(\mathbb{X} \in \{(x.y); x \le x_0, y \le y_0\}) = \alpha$ and $P(\{(x,y); x > x_0\}) = P(\{(x,y); y > y_0\})$. For simplicity write $A(\mathbb{X}, \mathbf{p}) = \inf_\alpha\{\mathbf{p} \in C(\mathbb{X}, \alpha)\}$. Then performance can be evaluated by estimating the probability $Q(\mathbb{X}) = P(\mathcal{R}_p(\mathbb{X}) \in C(\mathcal{R}_p(\mathbb{X}), A(\mathcal{R}_p(\mathbb{X}), \mathbf{p}))|\mathbb{X})$ using the simulated data for each of the five variations of generalized fiducial distribution. If the generalized fiducial inference were exact, $Q(\mathbb{X})$ should follow the $\mathrm{U}(0,1)$ distribution. The level of agreement of $Q(\mathbb{X})$ with a $\mathrm{U}(0,1)$ distribution was examined using QQ-plots.

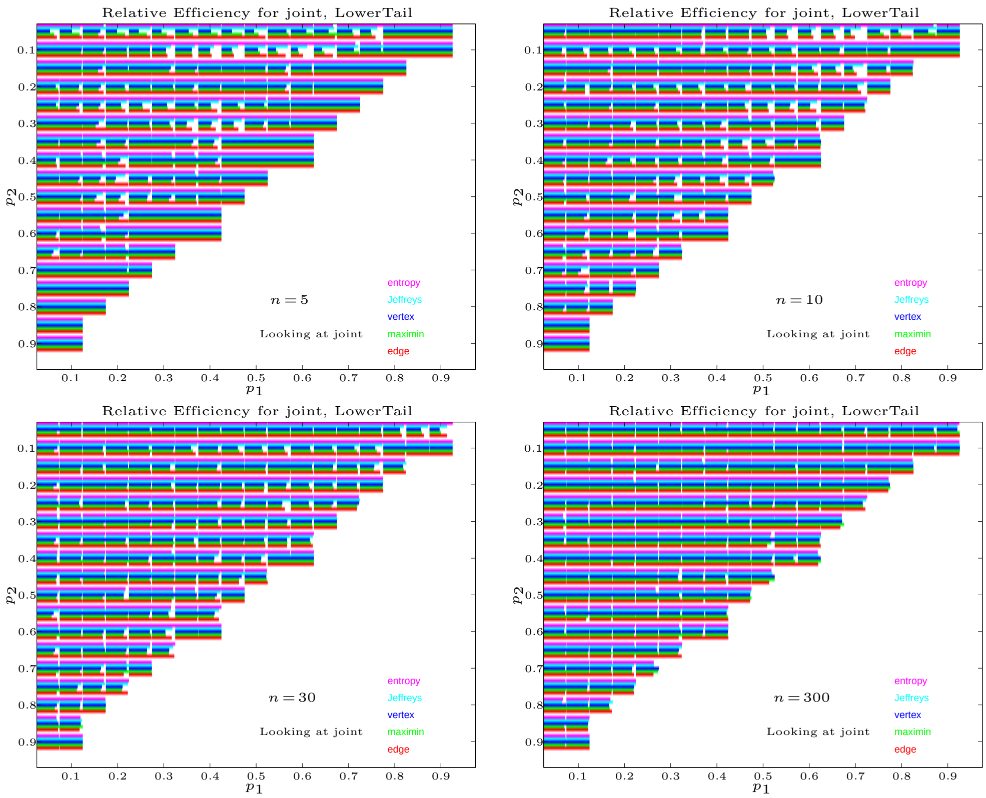

Figure 7. Plots of relative efficiency based on A_l for $n = 5, 10, 30, 300$. The longer the bar corresponding to a method, the better the method. The various colors correspond to choices of **D**.

As above, the function with values in the space of distributions $Q(\mathbb{X})$ takes only finitely many values. Thus it cannot be expected that the agreement with the uniform distribution is very good for small values of n. However, the agreement improves dramatically as n increases. Since the QQ-plots generated for the trinomial distribution are very similar to the figures shown in Example 6, we do not display them here.

Define A, A_l and A_u as in Example 6. Since we have one more parameter than in the binomial case, we need a new way to display the comparison between the procedures. For each fixed n, p_1, p_2 and each of the five procedures, we calculated a relative efficiency of procedure j as $\min_j A(j)/A(i)$, where $A(i)$ is the value of A for procedure i. Values close to 1 then mean a relatively good performance, while small values mean relatively poor performance.

For each fixed n we plotted an image containing a matrix of cells comparing these relative efficiencies. The cells were then placed on the image depending

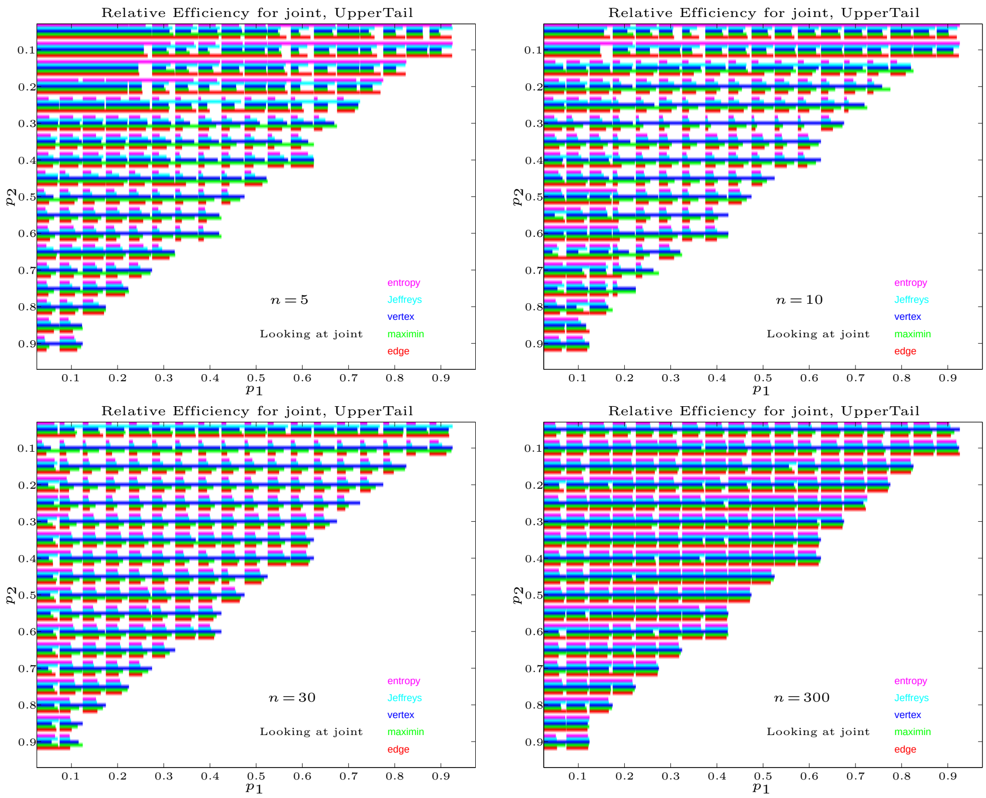

Figure 8. Plots of relative efficiency based on A_u for $n = 5, 10, 30, 300$. The longer the bar corresponding to a method, the better the method. The various colors correspond to choices of $\mathbf{D}$.

on values of p_1 and p_2. For illustration we show plots of these quantities for $n = 5, 10, 30, 300$ in Figures 7, 8 and 9.

The best choice for D is the first maximum variance choice (called Vertex in the figures). Notably, this choice seems to consistently outperform even the Bayesian posterior computed with respect to Jeffreys prior.

Generalized fiducial inference for a mixture of two normals.

Example 8. In this example we consider the generalized fiducial distribution for the parameters of a mixture of two normal distributions. This is a prototypical example that can be used to construct generalized fiducial distributions for many other problems. In particular, one can use these ideas to construct a robust generalized fiducial confidence interval for a mean of a normal sample by considering a mixture of normal and Cauchy distributions—see Glagovskiy (2006). To our knowledge this is the first time the fiducial paradigm has been used in such a

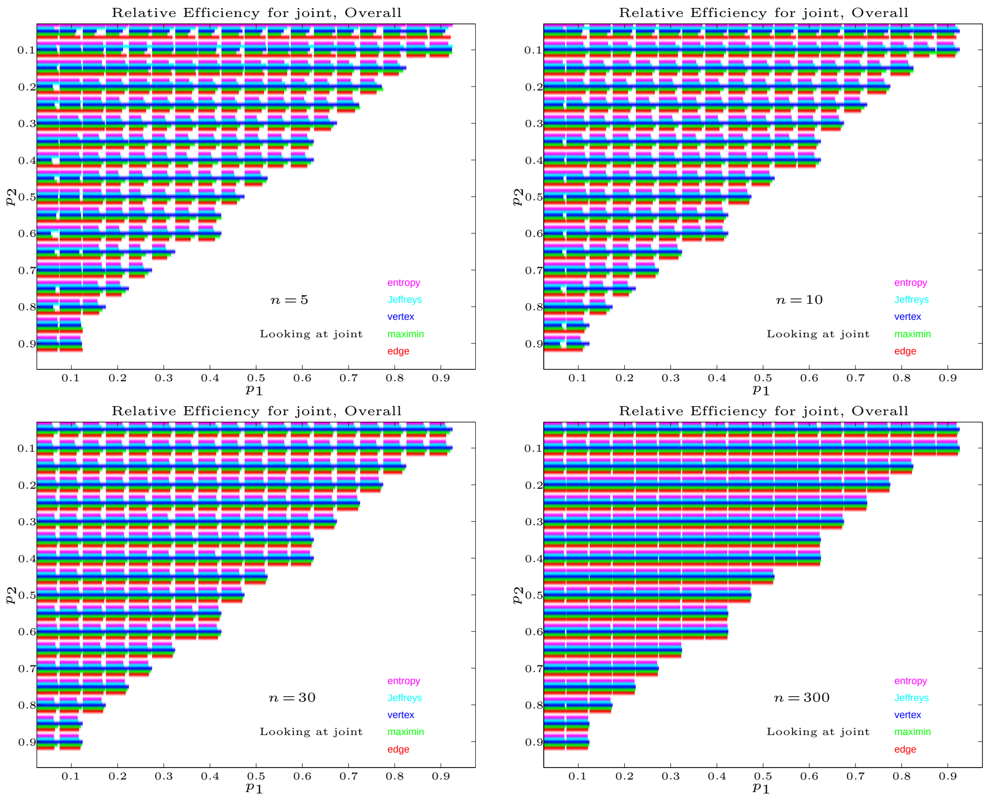

Figure 9. Plots of relative efficiency based on A for $n = 5, 10, 30, 300$. The longer the bar corresponding to a method, the better the method. The various colors correspond to choices of $\mathbf{D}$.

complex situation.

Let $X_1, \ldots, X_n$ be independent random variables that are either $\mathrm{N}(\mu_1, \sigma_1^2)$ or $\mathrm{N}(\mu_2, \sigma_2^2)$. Moreover, assume that each comes from the second distribution with probability p, independently of others. For identifiability reasons we assume that $\mu_1 < \mu_2$. We also assume that we observe at least two data points from each distribution. Our goal is to find the generalized fiducial distribution of $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p)$.

We can write a set of structural equations for $X_1, \ldots, X_n$ as

$$X_i = (\mu_1 + \sigma_1 Z_i) I_{(0,p)}(U_i) + (\mu_2 + \sigma_2 Z_i) I_{(p,1)}(U_i), \; i = 1, \ldots, n,$$

where Z_i are i.i.d. N(0,1) and U_i are i.i.d. U(0,1) random variables. When finding the set-valued function Q, we need to realize that this inversion will be stratified based on the possible assignment of the observed x_i to one of the two groups. For simplicity of notation the observed points x and the corresponding

z values assigned to Groups 1 and 2 are denoted by $v_1\ldots,v_k$ and $h_1,\ldots,h_k$, and $w_1,\ldots,w_{n-k}$ and $r_1,\ldots,r_{n-k}$ respectively. We can then write

$$
\begin{aligned}
&Q(x_1,\ldots,x_n;z_1,\ldots,z_n,u_1,\ldots u_n)\\
&=\begin{cases}
\left\{\left(\frac{h_1v_2-h_2v_1}{h_1-h_2},\left(\frac{v_1-v_2}{h_1-h_2}\right)^2,\frac{r_1w_2-r_2w_1}{r_1-r_2},\left(\frac{w_1-w_2}{r_1-r_2}\right)^2\right)\right\}\times(u_{s:n},u_{s+1:n}),\\
\quad\text{for each assignment of the } x_i \text{ to the two groups, if}\\
\quad v_l=\frac{h_1v_2-h_2v_1}{h_1-h_2}-\left|\frac{v_1-v_2}{h_1-h_2}\right|h_l,\ l=3,\ldots,s,\\
\quad\text{and}\quad w_l=\frac{r_1w_2-r_2w_1}{r_1-r_2}-\left|\frac{w_1-w_2}{r_1-r_2}\right|r_l,\ l=3,\ldots,n-s;\\
\emptyset\quad\text{otherwise.}
\end{cases}
\end{aligned}
$$

Similarly as in previous examples, for each possible assignment of the observations to the two groups, set $M_1=(H_1v_2-H_2v_1)/(H_1-H_2)$, $N_1=[(v_1-v_2)/(H_1-H_2)]^2$, $M_2=(R_1w_2-R_2w_1)/(R_1-R_2)$, $N_2=((w_1-w_2)/(R_1-R_2))^2$, $P=U_{s:n}+\bar{U}(U_{s+1:n}-U_{s:n})$, $K_l=(H_1v_2-H_2v_1)/(H_1-H_2)-|(v_1-v_2)/(H_1-H_2)|H_l$, $l=3,\ldots,s$, and $L_l=(R_1w_2-R_2w_1)/(R_1-R_2)-|(w_1-w_2)/(R_1-R_2)|R_l$, $l=3,\ldots,n-s$.

We interpret the conditional distribution (4.3) as

$$
\begin{aligned}
&\lim_{\varepsilon\to 0_+}\Bigg(\sum_{s=3}^{n-2}\sum_{\text{assignments}}P\Big(M_1\in(m_1,m_1+\varepsilon),N_1\in(n_1,n_1+\varepsilon),\\
&\qquad\qquad M_2\in(m_2,m_2+\varepsilon),N_2\in(n_2,n_2+\varepsilon),\\
&\qquad\qquad P\in(p,p+\varepsilon),K_l\in(v_l,v_l+\varepsilon),L_j\in(w_j,w_j+\varepsilon)\Big)\Bigg)\\
&\qquad\qquad\times\Bigg(\sum_{s=3}^{n-2}\sum_{\text{assignments}}P\Big(K_l\in(v_l,v_l+\varepsilon),L_j\in(w_j,w_j+\varepsilon)\Big)\Bigg)^{-1}\\
&=C^{-1}\sum_{s=3}^{n-2}\sum_{\text{assignments}}\frac{f_P(p,s)}{\binom{n}{s}}f_{M_1,N_1,\mathbf{K}}(m_1,n_1,\mathbf{v})f_{M_2,N_2,\mathbf{L}}(m_2,n_2,\mathbf{w}),
\end{aligned}
\tag{6.7}
$$

where f_P is as defined in (6.5) and both $f_{M_1,N_1,\mathbf{K}}$ and $f_{M_2,N_2,\mathbf{L}}$ are as defined in (4.5). The constant C on the left-hand-side of (6.7) is

$$
\begin{aligned}
C&=\sum_{s=3}^{n-2}\sum_{\text{assignments}}\int\cdots\int\frac{f_P(p,s)}{\binom{n}{s}}f_{M_1,N_1}(m_1,n_1,\mathbf{v})f_{M_2,N_2}(m_2,n_2,\mathbf{w}) &(6.8)\\
&=\sum_{s=3}^{n-2}\sum_{\text{assignments}}\frac{\Gamma(\frac{s-1}{2})\Gamma(\frac{n-s-1}{2})\frac{\sum_{1\le i<j\le s}|v_j-v_i|}{\binom{s}{2}}\frac{\sum_{1\le i<j\le n-s}|w_j-w_i|}{\binom{n-s}{2}}}{\binom{n}{s}4\pi^{\frac{n}{2}-1}\sqrt{s(n-s)}\left(\sum_{i=1}^{s}(v_i-\bar{v})^2\right)^{\frac{s-1}{2}}\left(\sum_{i=1}^{n-s}(w_i-\bar{w})^2\right)^{\frac{n-s-1}{2}}}
\end{aligned}
$$

Since the sums in the generalized fiducial distribution have a total of $2^n - 2 - 2n - n(n-1)$ terms, we are unable to get a closed form of the generalized fiducial density. However, we can still use the derived generalized fiducial distribution for inference by simulating observations from the generalized fiducial distribution using a Metropolis-Hastings algorithm.

The main idea is as follows. Once we know the assignment of observations to Groups 1 and 2, it is straightforward to generate the values of the 5-dimensional generalized fiducial distribution. This is done by calculating the corresponding sample means and variances for each group, and using (4.6) and (6.4).

To generate a random assignment, notice that each configuration assignment has a probability proportional to the corresponding summand in the right-hand-side of (6.8). We can therefore generate a proposal configuration by taking a previous assignment, randomly choosing one data point and switching it to the other group. This new proposed assignment is then rejected or accepted using the usual Metropolis-Hastings rule. Once we have a new random assignment, we then generate the observation from the 5-dimensional generalized fiducial distribution. The stationary distribution of the assignment-valued Markov chain is clearly the generalized fiducial distribution of the assignment. Therefore this procedure generates observations from the generalized fiducial distribution after an adequate burn-in period. It is worth pointing out that even though this procedure is computationally intensive, it is usable for most situations encountered in practice.

To evaluate performance, we conducted a small scale simulation study. We considered a mixture of the $N(-1, 1/27)$ and $N(0, 9)$ distributions, with $n = 80$, $n = 250$ and the mixing proportion $p = 0.65$. We also considered $N(-1.5, 1)$ and $N(1.5, 1)$ with $n = 100, 250$ and $p = 0.6$. We wish to remark that the second mixture is actually very hard to estimate. We used the particular choice $D \sim$ Beta$(1/2, 1/2)$ in the definition of f_P in (6.7), cf. (6.5).

For each model we generated a sample from the generalized fiducial distribution and used it to find a sample from $Q_1(\mathbb{X}) = P(\mathcal{R}_{\mu_1}(\mathbb{X}^\star) < \mu_1|\mathbb{X})$, $Q_2(\mathbb{X}) = P(\mathcal{R}_{\mu_2}(\mathbb{X}^\star) < \mu_2|\mathbb{X})$, $Q_d(\mathbb{X}) = P(\mathcal{R}_{\mu_2-\mu_1}(\mathbb{X}^\star) < \mu_2 - \mu_1|\mathbb{X})$, and $Q_p(\mathbb{X}) = P(\mathcal{R}_p(\mathbb{X}^\star) < p|\mathbb{X})$. Notice that Q_d is measuring the performance of a fiducial solution for a generalization of a Beherns-Fisher problem where we want a CI for $\mu_2 - \mu_1$ but do not know which observations belong to which group.

If the inference based on generalized fiducial distributions were exact, these random variables would follow a U$(0, 1)$ distribution. To check for agreement, we constructed QQ-plots. These can be found in Figures 10, 11, 12 and 13. We see that, while the agreement is not very good in the body of the distribution, it is actually very good in the tails. This means that the inference based on the

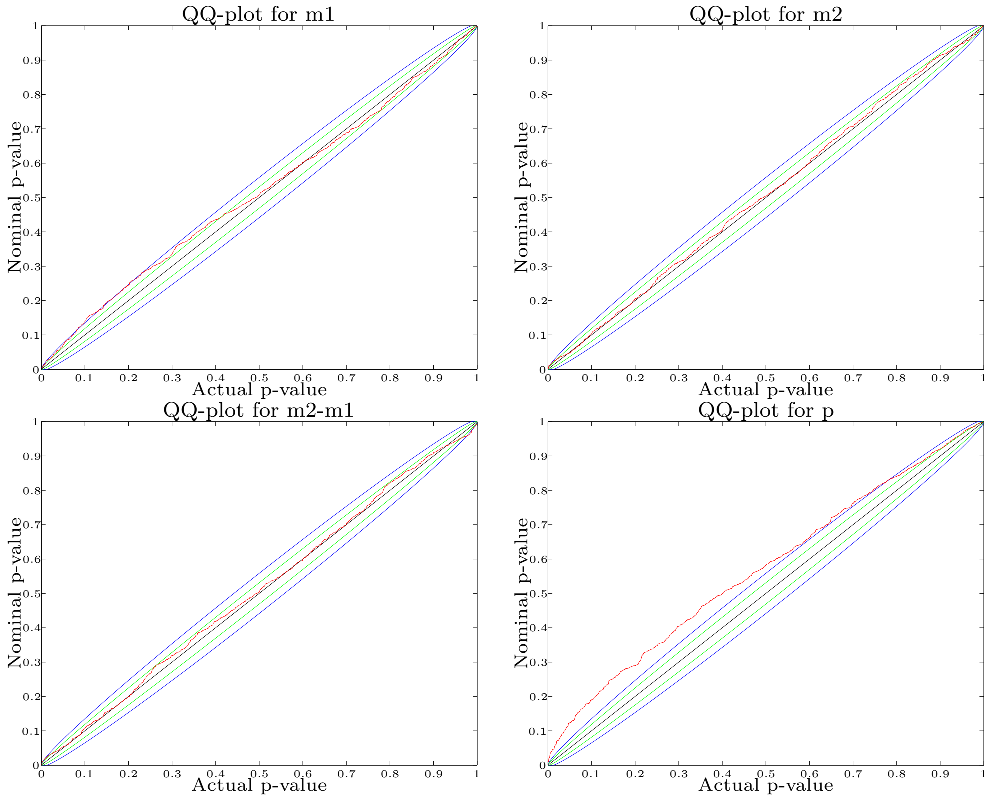

Figure 10. QQ-plots of $Q_1(\mathbb{X}), Q_2(\mathbb{X}), Q_d(\mathbb{X})$ and $Q_p(\mathbb{X})$ for $n = 80$ observations of the $p = 0.65$ mixture of $N(-1, 1/27)$ and $N(0, 9)$. The blue and green envelope correspond to an area of natural fluctuation of a QQ-plot due to randomness taken uniformly and pointwise, respectively. The QQ-plot is based on 1,000 replications.

generalized fiducial distribution has approximately the stated coverage. We also see that the inference for μ_1, μ_2 and $\mu_2 - \mu_1$ seems more accurate than for p, which is often too conservative. In any case, the performance seemed very good given the fact we chose mixtures that are hard to estimate. Finally, we remark that the fit improves for larger n. This leads us to conjecture that the inference will be correct asymptotically as $n \to \infty$.

Multiple Comparison

Example 9. We include this last example to show that the regions defined in Assumptions 1.3 are flexible enough to allow for typical multiple comparison intervals.

Suppose that for each $i = 1, \ldots, K$, Y_{ij}, $j = 1, \ldots, n_i$, is i.i.d. $\mathrm{N}(\mu_i, \sigma_i^2)$.

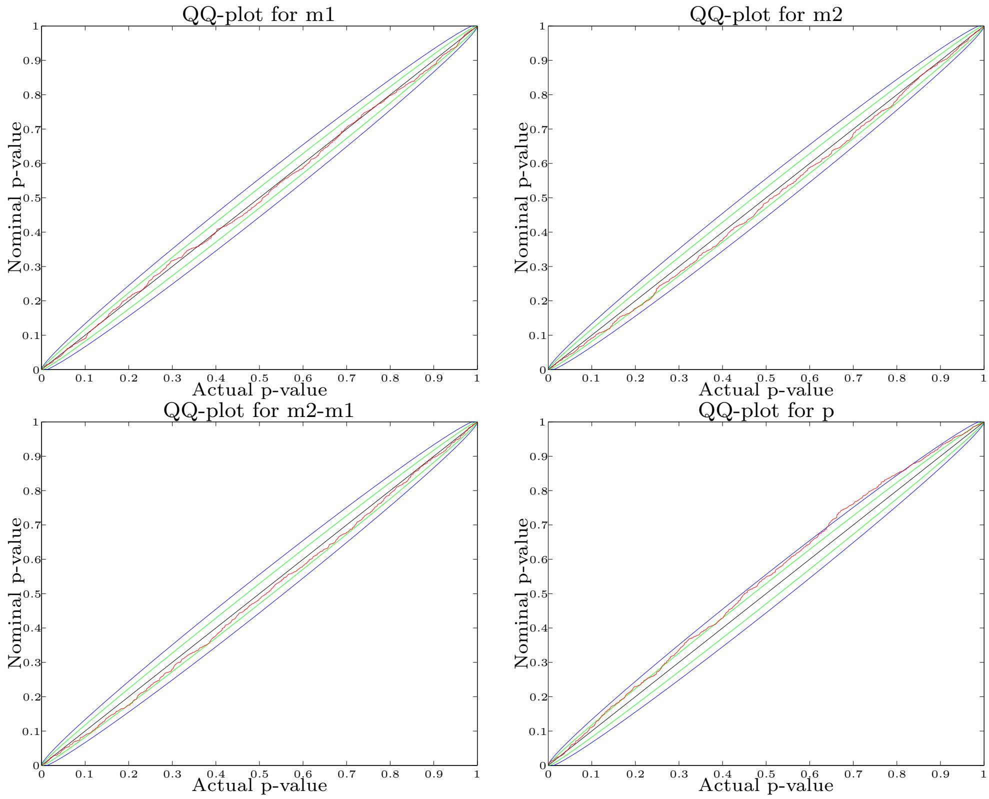

Figure 11. QQ-plots of $Q_1(\mathbb{X}), Q_2(\mathbb{X}), Q_d(\mathbb{X})$ and $Q_p(\mathbb{X})$ for $n = 250$ observations of the $p = 0.65$ mixture of $N(-1, 1/27)$ and $N(0, 9)$. The blue and green envelope correspond to an area of natural fluctuation of a QQ-plot due to randomness taken uniformly and pointwise, respectively. The QQ-plot is based on 1,105 replications.

The K samples are assumed independent of each other. We are interested in the problem of constructing simultaneous confidence intervals for $\delta_{ij} = \mu_i - \mu_j$ for all $i \neq j$.

We first observe that, by independence, the generalized fiducial distribution for δ_{ij} is the same as the distribution of the GFQ given by $\mathcal{R}_{\delta_{ij}}(\mathbb{S}) = \mathcal{R}_{\mu_i} - \mathcal{R}_{\mu_j}$, where the GPQ for μ_p is

$$\mathcal{R}_{\mu_p} = \bar{Y}_p - \frac{S_p T_p^\star}{\sqrt{n_p}}$$

and the $T_p^\star \sim t_{n_p - 1}$ are independent of the data and of each other.

Define

$$\mathcal{D}(\mathbb{S}) = \max_{i \neq j} \left| \frac{(\overline{Y}_i - \overline{Y}_j) - \mathcal{R}_{\delta_{ij}}(\mathbb{S})}{\sqrt{V_{ij}}} \right|,$$

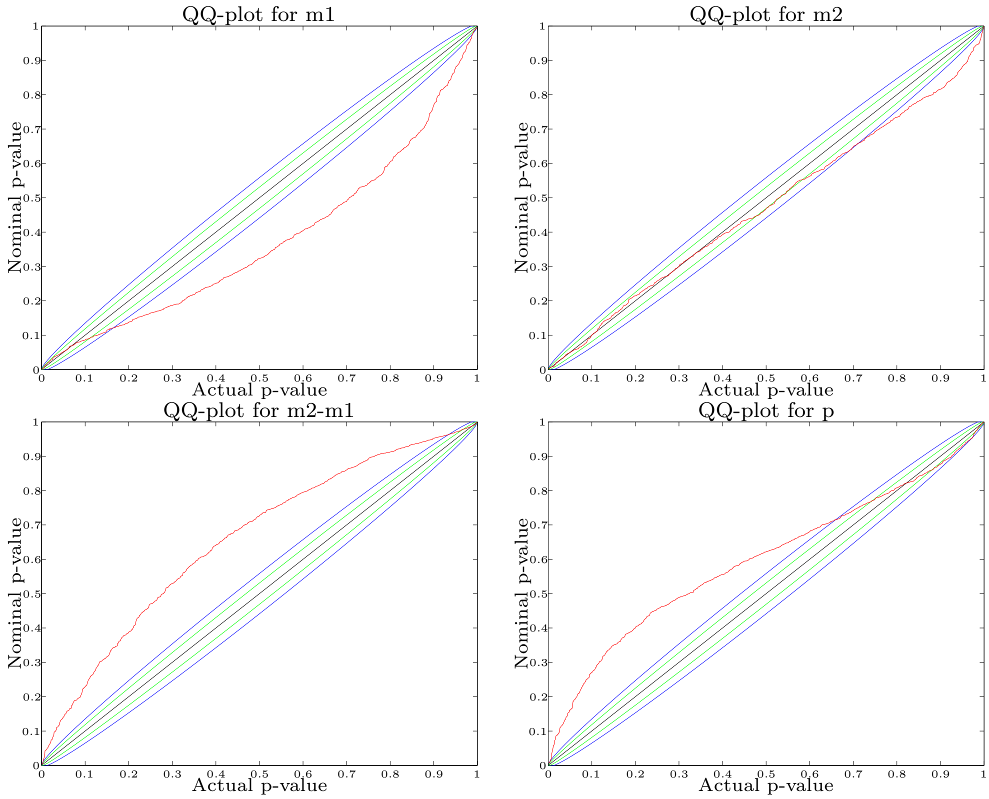

Figure 12. QQ-plots of $Q_1(\mathbb{X}), Q_2(\mathbb{X}), Q_d(\mathbb{X})$ and $Q_p(\mathbb{X})$ for $n = 100$ observations of the $p = 0.6$ mixture of $N(-1.5, 1)$ and $N(1.5, 1)$. The blue and green envelope correspond to an area of natural fluctuation of a QQ-plot due to randomness taken uniformly and pointwise, respectively. The QQ-plot is based on 1,000 replications.

where $V_{ij} = \overline{Y}_i - \overline{Y}_j$, i.e., $V_{ij} = \frac{S_i^2}{n_i} + \frac{S_j^2}{n_j}$. The $100(1-\alpha)\%$ two-sided simultaneous generalized fiducial CIs for pairwise differences $\delta_{ij}, i \neq j$, of means of more than two independent normal distributions are $[L_{ij}, U_{ij}]$ where

$$\begin{aligned} L_{ij} &= \overline{Y}_i - \overline{Y}_j - d_{1-\alpha}\sqrt{V_{ij}}, \\ U_{ij} &= \overline{Y}_i - \overline{Y}_j + d_{1-\alpha}\sqrt{V_{ij}} \end{aligned} \tag{6.9}$$

and d_γ denotes the 100γ-percentile of the conditional distribution of $\mathcal{D}(\mathbb{S}, \mathbb{S}^*, \xi)$ given $\mathbb{S} = \mathbf{s}$.

To set up confidence regions one can use simulation. The simultaneous generalized fiducial confidence intervals for δ_{ij} do not lead to exact frequentist inference. However, simulation results suggest very good practical properties.

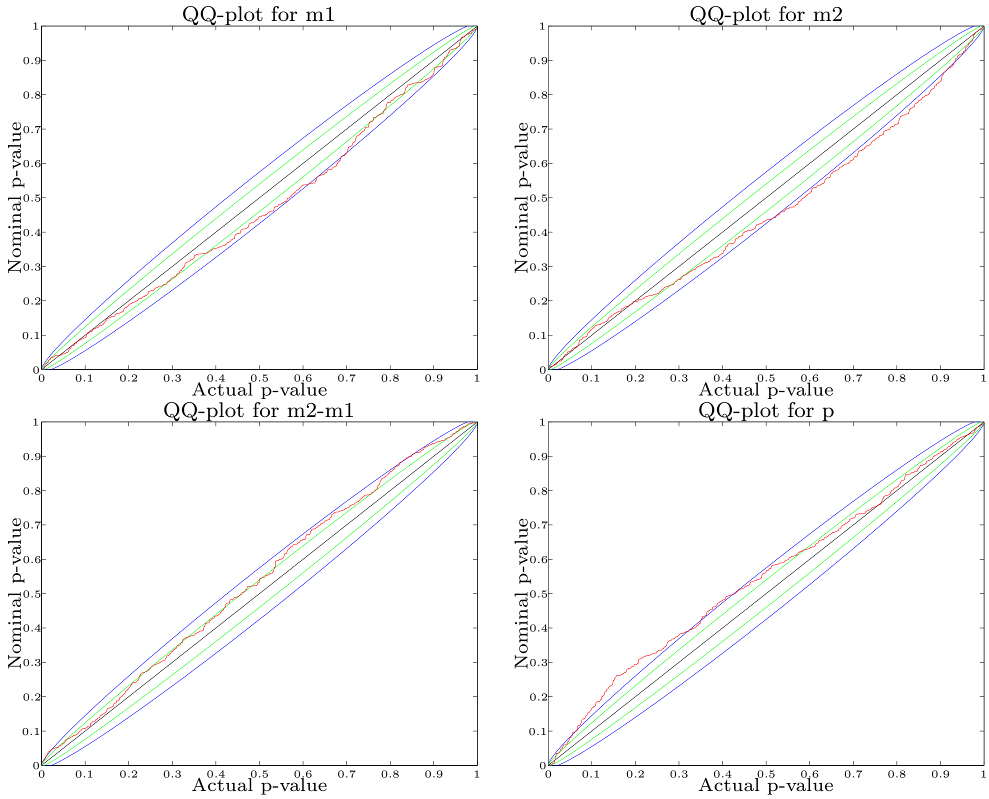

Figure 13. QQ-plots of $Q_1(\mathbb{X}), Q_2(\mathbb{X}), Q_d(\mathbb{X})$ and $Q_p(\mathbb{X})$ for $n = 500$ observations of the $p = 0.6$ mixture of $N(-1.5, 1)$ and $N(1.5, 1)$. The blue and green envelope correspond to an area of natural fluctuation of a QQ-plot due to randomness taken uniformly and pointwise, respectively. The QQ-plot is based on 600 replications.

For details on the simulation and some generalization we refer the reader to Abdel-Karim (2005) and Hannig, E, Abdel-Karim and Iyer (2006a).

To show that the generalized fiducial distribution leads at least to asymptotically proper frequentist coverage, define $n = \sum_{k=1}^{K} n_k$ and assume that $n_i/n \to p_i \in (0, 1)$. It is fairly straightforward to see that $\mathbb{S} = (\overline{Y}_1, S_1^2, \ldots, \overline{Y}_K, S_K^2)^\top$ satisfies Assumption 1.1. Similarly, $\mathcal{R} = (\mathcal{R}_{\delta_{12}}, \mathcal{R}_{\delta_{13}} \ldots, \mathcal{R}_{\delta_{(K-1)K}})^\top$ satisfies Assumptions 1.2 with the $K(K-1)/2 \times 2K$ matrix

$$A = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & -1 & 0 \end{pmatrix}.$$

Similarly, the assumption in (5.2) is satisfied with the function $\zeta(\mathbb{S}) = A \cdot \mathbb{S}$

Finally, we need to show that the region described in (6.9) satisfies Assumptions 1.3. To that end, observe that the conditional distribution of $\mathcal{D}(\mathbb{S})|\mathbb{S}$ could be represented as function of distribution of $\mathcal{R}$, $\zeta(\mathbb{S})$, and $\mathbb{S}$. Here the estimator of variance nV_{ij} is a continuous function of $\mathbb{S}$. The various conditions of this assumption now follow by Slutsky's Lemma and simple algebra.

7. Non-Uniqueness of Fiducial Distribution

The generalized fiducial recipe of Section 4 seems to provide an approach for deriving statistical procedures that have good properties. Unfortunately, it does not lead to a unique generalized fiducial distribution. There are three main sources of non-uniqueness. First, non-uniqueness is due to the choice of the structural equation (4.1). However, even if we decide to fix the structural equation and make it a part of the model, generalized fiducial distribution is still not defined uniquely.

Out of the two remaining sources of non-uniqueness the more obvious one is the fact that the sets $Q(\mathbf{X}, U^*)$ might have more than one element. This means that we would not be able to find the exact value of ξ even if we knew both $\mathbb{X}$ and U. Consequently, the data itself is not able to tell us which value of ξ was used. In order to resolve this non-uniqueness one has to have some *apriori* way of choosing between the elements of $Q(\mathbf{X}, U^*)$. Fortunately, in many application we observe that $\sqrt{n}\,\mathrm{diam}(Q(\mathbf{X}, U^*)) \to 0$. This means that in these cases the role of the *apriori* information is asymptotically negligible. In other words the uncertainty in the fiducial distribution comes mainly from the uncertainty in the distribution of $Q(\mathbf{X}, U^*)|\{Q(\mathbf{X}, U^*) \neq \emptyset\}$, which involves no subjective choice and is typically of the order of $n^{-1/2}$. Only a small portion of the uncertainty in the fiducial distribution comes from $V(\bullet)$; this is subjective and usually has order of n^{-1}. This is in contrast to Bayesian inference where the prior influences the whole posterior distribution and not just a part of it. Of course such a decomposition can be expected only in parametric problems and, just as with the choice of a prior in Bayesian methods, the *apriori* choice of $V(Q(\mathbf{x}, u))$ plays a big role in non-parametric and semi-parametric problems.

Based on our experience with the problems we investigated, we recommend the use of a $V(Q(\mathbf{x}, u))$ that is independent of the data and that maximizes the determinant of the variance of the generalized fiducial distribution. Another useful option is to use the uniform distribution on $Q(\mathbf{x}, u)$. This second option should work reasonably well and be fairly easy to implement even if we deal with higher-dimensional problems.

The final source of non-uniqueness is the Borel paradox. If in the generalized fiducial recipe (4.3) we have $P(Q(\mathbf{x}, u) \neq \emptyset) = 0$, the resulting generalized fiducial distribution depends on the way we decide to interpret the conditioning. We consider this to be a more severe problem because it is much harder to investigate and resolve. To demonstrate the severity of the situation, consider the following continuation of Example 3.

Example 10. Let $X_1,\dots,X_n$ be i.i.d. $\mathrm{N}(\mu,\sigma^2)$. In Example 3 we showed two different way of implementing the generalized fiducial recipe that led to the same desirable solution. Unfortunately, there are many other ways of implementing the generalized fiducial recipe that do not lead to good solutions. We demonstrate one of them here.

We again write the structural equation as

$$X_i = \mu + \sigma Z_i,\ \ i = 1,\dots,n.$$

For simplicity of notation take $n = 2k$. If

$$M_j = \frac{z_{2j-1}x_{2j} - z_{2j}x_{2j-1}}{z_{2j-1} - z_{2j}}, \quad H_j = \left(\frac{x_{2j-1} - x_{2j}}{z_{2j-1} - z_{2j}}\right)^2 \quad j = 1,\dots,k,$$

we can write

$$Q(x_1,\dots,x_n;z_1,\dots,z_n) = \begin{cases} \{(M_1,H_1)\} & \\ \qquad \text{if } M_j = M_1,\ H_j = H_1,\ j = 2,\dots,k \\ \emptyset \qquad \text{otherwise.} \end{cases}$$

Defining $D_{j,1} = M_j - M_1, \quad D_{j,2} = H_j - H_1,\ j = 2,\dots k$, we can interpret the generalized fiducial distribution (4.3) as the conditional distribution of $(M_1,H_1)|\mathbf{D} = \mathbf{0}$. A simple calculation shows that this conditional distribution has density

$$f_{\mathcal{R}_{(\mu,\sigma^2)}}(m,h) = \frac{e^{-(m-\bar{x}_n)^2/(2h/n)-(n-1)s_n^2/(2h)}((n-1)s_n^2)^{n-3/2}}{\sqrt{\pi/n}\,\Gamma\,(n-3/2)\,2^{n-1}h^n} I_{(0,\infty)}(h). \quad (7.1)$$

Here $\bar{x}_n = \sum_{i=1}^n x_i/n$ and $s_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2/(n-1)$. The distribution derived in (7.1) is different from the one derived in (4.6). In fact inference based on (7.1) will not lead to correct frequentist inference. In fact the coverage probability of any lower tail confidence interval converges to 0 as $n \to \infty$.

The problem illustrated in Examples 3 and 10 is an instance of Borel paradox — see for example Section 4.9.3 of Casella and Berger (2002), and also Hannig (1996), for a thorough discussion of this paradox. The main message of the Borel paradox is that conditioning on an event of probability zero greatly depends on the context in which we interpret the conditions.

Consider in particular $X|Y = 0$, where (X,Y) is jointly continuous. There is a random variable U such that (X,U) is jointly continuous and $\{Y = 0\} =$

$\{U = 0\}$, but the conditional density of $X|Y = 0$ is different from the conditional density $X|U = 0$. Since there is no theoretical reason to deem either $X|Y = 0$ or $X|U = 0$ superior, people often rely on the context of the problem to make the choice, e.g., conditional distributions in regression settings. However, one can often come up with a modification of the "story" behind the problem that leads naturally to a different choice of the conditioning variable. This can be then presented as a paradox-two apparently equivalent formulations of the same statistical problem lead to different answers.

Remark 7. Fraser (1968) has linked fiducial inference with group structure. A very good explanation of his ideas can be found in Appendix 3 of Dawid et al. (1973). Fraser's assumption of group structure can guide one in the choice of structural equation. In particular we can choose $\mathbf{X} = g\mathbf{U}$, where $g \in G$ is an element of a group acting on the random vector $\mathbf{U}$. Additionally, with this choice of structural equation the set $Q(\mathbf{x}, \mathbf{u})$ is trivially guaranteed to have at most 1 element for all choices of $\mathbf{x}$ and $\mathbf{u}$. Thus, some sources of non-uniqueness are eliminated. Unfortunately, the second source of non-uniqueness, Borel paradox, is still present. We again need to interpret a conditional probability that is conditioned on an event that has probability 0. Having the group structure presents us with a natural choice of conditioning σ-algebra, the maximal invariant σ-algebra $\mathcal{I}$. Unfortunately, the problem of non-uniqueness is still present as demonstrated by Example 7 in Dawid et al. (1973), where the authors show that addition of information clearly irrelevant to the inference leads to a different fiducial distribution. This could be explained again by a phenomenon related to the Borel paradox. The paradox of Dawid et al. (1973) is based on the fact that the following should be true: $P(A|B_1) = P(A|B_2) = \cdots$ implies $P(A|\bigcup B_i) = P(A|B_1)$. This is in fact true if $P(\bigcup B_i) > 0$. It could fail otherwise because of the Borel paradox. Dawid et al. (1973) have their conditioning sets B_i depend on the value of an added parameter λ. By symmetry in their example, all conditional probabilities of interest should be the same. What happens is that depending on whether the value of λ is fixed (conditioning on B_i) or is unknown (conditioning on $\bigcup B_i$) changes the natural σ-algebra for conditioning, the invariant σ-algebra. This leads to a different answer in each case, hence the paradox.

Finally we remark that if the group G is sufficiently complicated, computing the conditional expectation with respect to $\mathcal{I}$ could be quite complicated. Therefore, we still suggest using (4.11) whenever applicable, as it can be computed relatively easily.

Remark 8. One particular way of avoiding the Borel paradox presents itself in the case when the parameter space is an open set in $\mathbb{R}^p$ and the model allows for a p-dimensional complete sufficient statistic that is a smooth function of the data. In this case we can first reduce the data by obtaining complete sufficient statistics and then applying the generalized fiducial recipe to the distribution of the complete sufficient statistics. A simple Jacobian calculation shows that the generalized fiducial distribution is independent of the particular form of the complete sufficient statistics we used. This idea has been used in the first part of Example 3.

8. Conclusions

In this paper we studied the properties of generalized fiducial distributions without relying on any additional group assumptions. We have shown how the fiducial argument could be applied to several problems, and demonstrated by simulation that it leads to statistical procedures with good small sample frequentist properties. We also investigated the asymptotic properties of generalized fiducial distributions and showed that in many examples a generalized fiducial distribution has good asymptotic properties. Thus fiducial inference appears to be a good tool for deriving statistical procedures and should not be ignored by the statistical community.

Finally we investigated an inherent non-uniqueness of fiducial inference that is in some way similar to the non-uniqueness of Bayesian inference due to the choice of a prior. We argued that the non-uniqueness of fiducial inference is essentially caused by the Borel paradox, the fact that the conditional distribution conditioned on an event of probability 0 is not uniquely determined. In fact, in our opinion, the Borel paradox is the root cause for most of the paradoxes associated with fiducial inference. The future of the fiducial argument can be driven by new and exciting applications. The most promising from this point of view seems to be the fiducial distribution for a mixture of two normals in Example 8. In fact the ideas of that example are extended in a current work of Hannig and Lee (2007) that uses it for wavelet thresholding with promising results. Another possible use is the detection of significant p-values in microarray experiments where a mixture of uniform and beta distribution is typically used Allison, Gadbury, Heo, Fernández, Lee, Prolla and Weindruch (2002) and Pounds and Morris (2003).

An important step in the practical application of generalized fiducial inference would be finding a simple workable formula. We have made a first step in this direction at (4.11). However, a better understanding of the issues surrounding non-uniqueness is needed. A possible tool that can guide our choices can be the study of higher order asymptotic properties of generalized fiducial distribution; another tool is inspired by ideas of Dempster (2008). Due to limitations on precision of measuring devices we never observe a continuous random variable.

This can be used to change (4.3) so that we always condition on an event of positive probability at the expense of having a bigger non-uniquness due to the choice of $V(\bullet)$. Even though this idea is very appealing, more research is needed to fully understand the computational issues invlolved.

The surprisingly good small sample properties demonstrated in many statistical applications lead us to believe that if computer simulations had been available 60 years ago, fiducial argument would have been part of statistical mainstream today. We hope that this paper will stimulate discussion, further development, and more appreciation for the great minds who have worked on this topic in the past.

Acknowledgements

I would like to thank Prof. Hari Iyer, who introduced me to the field of fiducial inference, for many useful conversations that were critical in developing this manuscript. I would also like to thank Yuriy Glagovskiy, Prof. Paul Mielke, and Prof. Thomas Lee for finding numerous typos in an earlier version of this manuscript. I also want to thank the Statistics Department at The University of Chicago for providing hospitality to the author during work on a revised version of this paper. Jan Hannig's research is supported in part by National Science Foundation under grants DMS-0504737 and DMS-0707037.

Appendix. Proofs

Proof. of Theorem 1. By Assumption 1 and Skorokhod's Representation Theorem Billingsley (1995) imply that we can assume without loss of generality that

$$\sqrt{n}\,(\mathbb{S}-\mathbf{t}(\xi)) \to \mathbf{H} \ a.s.. \tag{A.1}$$

This, Assumptions 2a, and (5.2) ensure that

$$\sqrt{n}(\mathcal{R}_\theta(\mathbb{S})-\theta) \xrightarrow{\mathcal{D}} R(\mathbf{H}) \ a.s. \quad \sqrt{n}(\zeta(\mathbb{S})-\theta) \to A\mathbf{H} \ a.s.. \tag{A.2}$$

(Here the a.s. means for almost all sample paths of the process $\mathbb{S}_n$, and subsequently almost all values of $\mathbf{H}$.) Therefore by (A.1), (A.2), and Assumption 3d

$$C\left(\sqrt{n}(\mathcal{R}_\theta(\mathbb{S})-\theta), \sqrt{n}(\zeta(\mathbb{S})-\theta), \mathbb{S}, \gamma_n\right) \to C\left(R(\mathbf{H}), A\mathbf{H}, \mathbf{t}(\xi), \gamma\right) \ a.s.. \tag{A.3}$$

Also, by Assumption 3c we see that

$$P_\xi\left(\theta \in C(\mathcal{R}_\theta(\mathbb{S}), \zeta(\mathbb{S}), \mathbb{S}, \gamma_n)\right) = P_\xi\left(0 \in C(\sqrt{n}(\mathcal{R}_\theta(\mathbb{S})-\theta), \sqrt{n}(\zeta(\mathbb{S})-\theta), \mathbb{S}, \gamma_n)\right).$$

To finish the proof, we show that

$$P_\xi(0 \in C(\sqrt{n}(\mathcal{R}_\theta(\mathbb{S})-\theta), \sqrt{n}(\zeta(\mathbb{S})-\theta), \mathbb{S}, \gamma_n)) \to P_\xi(0 \in C\left(R(\mathbf{H}), A\mathbf{H}, \mathbf{t}(\xi), \gamma\right)).$$

First notice that $R(\mathbf{h})-A\mathbf{h}$ has a multivariate normal distribution with mean zero and covariance matrix Σ_R. It is the same distribution as the distribution of $-A\mathbf{H}$. Assumption 2a implies

$$\begin{aligned}\{\mathbf{h} : 0 \in C(R(\mathbf{h}), A\mathbf{h}, \mathbf{t}(\xi), \gamma))\} &= \{\mathbf{h} : -A\mathbf{h} \in C(R(\mathbf{h}) - A\mathbf{h}, 0, \mathbf{t}(\xi), \gamma)\} \\ &= \{\mathbf{h} : -A\mathbf{h} \in C(-A\mathbf{H}, 0, \mathbf{t}(\xi), \gamma)\}. \qquad \text{(A.4)}\end{aligned}$$

For simplicity of notation, write $\mathbf{H}_n = \sqrt{n}(\mathbb{S} - \mathbf{t}(\xi))$. Also let

$$B_n=\{\mathbf{h} : 0\in C(\sqrt{n}(\mathcal{R}_\theta(\mathbf{t}(\xi) + \frac{\mathbf{h}}{\sqrt{n}})-\theta), \sqrt{n}(\zeta(\mathbf{t}(\xi) + \frac{\mathbf{h}}{\sqrt{n}})-\theta), (\mathbf{t}(\xi)+\frac{\mathbf{h}}{\sqrt{n}}), \gamma_n)\}$$

and $B = \{\mathbf{h} : 0 \in C(R(\mathbf{h}), A\mathbf{h}, \mathbf{t}(\xi), \gamma)\}$. The sets are chosen to satisfy

$$\begin{aligned}\{0 \in C(\sqrt{n}(\mathcal{R}_\theta(\mathbb{S}) - \theta), \sqrt{n}(\zeta(\mathbb{S}) - \theta), \mathbb{S}, \gamma_n)\} &= \{\mathbf{H}_n \in B_n\}, \\ \{0 \in C(R(\mathbf{H}), A\mathbf{H}, \mathbf{t}(\xi), \gamma)\} &= \{\mathbf{H} \in B\}.\end{aligned}$$

As noted before we have $\mathbf{H}_n \xrightarrow{\mathcal{D}} \mathbf{H}$. Moreover Assumptions 2b and 3d imply that B is open, $\partial B = \{\mathbf{h} : 0 \in \partial C(R_\theta(\mathbf{h}), \gamma)\}$, and $B_n \to B$. Assumption 3a and (A.4) additionally imply that $P(\mathbf{H} \in \partial B) = 0$.

Let $D_m = \overline{\bigcup_{k=m}^\infty B_k} \setminus (\bigcap_{k=m}^\infty B_k)^\circ$. Notice that by Assumption 3d we have $D_m \downarrow D \subset \partial B$ and $P(H \in D) = 0$. Moreover, if $m \le n$, $B_n \triangle B \subset D_m$.

Fix an $\varepsilon > 0$. Continuity of probability implies that there is an m_1 such that $P_\xi(\mathbf{H} \in C_{m_1}) < \varepsilon$. Consequently, convergence in distribution implies that there is an m_2 such that, for all $n > m_2$, $P_\xi(\mathbf{H}_n \in C_{m_2}) < \varepsilon$. This implies that for $n > \max(m_1, m_2)$,

$$|P_\xi(\mathbf{H}_n \in B_n) - P(\mathbf{H}_n \in B)| \le P(\mathbf{H}_n \in C_{m_1}) < \varepsilon.$$

Finally notice that

$$\begin{aligned}&|P_\xi(\mathbf{H}_n \in B_n) - P_\xi(\mathbf{H} \in B)| \\ &\quad \le |P_\xi(\mathbf{H}_n \in B_n) - P_\xi(\mathbf{H}_n \in b)| + |P_\xi(\mathbf{H} \in B_n) - P_\xi(\mathbf{H} \in B)|.\end{aligned}$$

Thus the Assumption 3b and (A.4), together with the definition of convergence in distribution, imply

$$\begin{aligned}P_\xi(0 \in C(\sqrt{n}(\mathcal{R}_\theta(\mathbb{S}) - \theta), \sqrt{n}(\zeta(\mathbb{S}) - \theta), \mathbb{S}, \gamma_n) &= P_\xi(\mathbf{H}_n \in B_n) \\ &\to P_\xi(\mathbf{H} \in B) = P_\xi(0 \in C\,(R(\mathbf{H}), A\mathbf{H}, \mathbf{t}(\xi), \gamma) = \gamma.\end{aligned}$$

This concludes the proof of the theorem.

Proof of Theorem 2. Assume without loss of generality that $\Theta = \mathbb{R}$, and recall the definition of $J(\mathbf{x}, \theta)$ from (4.13). To emphasize its dependence on n

we write $J_n(\mathbf{x},\theta)$ in the rest of the proof. The Strong Law of Large Numbers implies for all θ, $J_n(\mathbb{X},\theta) \to \pi(\theta), P_{\theta_0}$-$a.s.$. By the Dominated Convergence Theorem, $\pi(\theta)$ is continuous and positive on a neighborhood of θ_0. Moreover, since $J(\bullet,\theta)$ is measurable in x we have by the Uniform Strong Law of Large Numbers (Ghosh and Ramamoorthi (2003, Thm. 1.3.3)),

$$\sup_{\theta\in(\theta_0-\delta_0,\theta_0+\delta_0)} |J_n(\mathbb{X},\theta)-\pi(\theta)| \to 0,\ P_{\theta_0}\text{-}a.s.. \tag{A.5}$$

We closely follow the proof of of the Bernstein-von Mises theorem stated in the Theorem 1.4.2 of Ghosh et al. (2003). We can write

$$\pi_r(\theta,\mathbf{x}) = \frac{J_n(\mathbf{x},S_n+\theta/\sqrt{n})e^{L_n(S_n+\theta/\sqrt{n})-L_n(S_n)}}{\int_{\mathbb{R}} J_n(\mathbf{x},S_n+\theta'/\sqrt{n})e^{L_n(S_n+\theta'/\sqrt{n})-L_n(S_n)}\,d\theta'}.$$

Just as Ghosh et al. (2003), we first prove that

$$\int_{\mathbb{R}} \left| J_n(\mathbf{x},S_n+\frac{\theta'}{\sqrt{n}})e^{L_n(S_n+\frac{\theta'}{\sqrt{n}})-L_n(S_n)} - \pi(\theta_0)e^{-\frac{\theta' I(\theta_0)}{2}} \right| d\theta' \xrightarrow{P_{\theta_0}} 0. \tag{A.6}$$

Given any $0<\delta<\delta_0$ and $c>0$, we break $\mathbb{R}$ into three regions:

$$A_1=\{t:\ |t|<c\log\sqrt{n}\},\ A_2=\{t:\ c\log\sqrt{n}<|t|<\delta\sqrt{n}\},$$

and $$A_3=\{t:\ |t|>\frac{\delta\sqrt{n}}{2}\}.$$

On $A_1\cup A_2$ we compute

$$\begin{aligned}\int_{A_1\cup A_2} &\left|J_n(\mathbf{x},S_n+\frac{t}{\sqrt{n}})e^{L_n(S_n+\frac{t}{\sqrt{n}})-L_n(S_n)} - \pi(\theta_0)e^{-\frac{tI(\theta_0)}{2}}\right|dt \\ &\le \int_{A_1\cup A_2}\left|J_n(\mathbf{x},S_n+\frac{t}{\sqrt{n}})-\pi(S_n+\frac{t}{\sqrt{n}})\right| e^{L_n(S_n+\frac{t}{\sqrt{n}})-L_n(S_n)}dt \\ &\quad + \int_{A_1\cup A_2}\left|\pi(S_n+\frac{t}{\sqrt{n}})e^{L_n(S_n+\frac{t}{\sqrt{n}})-L_n(S_n)} - \pi(\theta_0)e^{-\frac{tI(\theta_0)}{2}}\right|dt. \end{aligned} \tag{A.7}$$

Since $\pi(\theta)$ (possibly truncated at large values) is a prior, the fact that the second term on the right-hand side of (A.7) goes to 0 in probability follows from the Bayesian version of the Bersntein-von Mises theorem. For details, see the proof of Theorem 1.4.3 in Ghosh et al. (2003).

Notice that

$$\begin{aligned}\int_{A_1\cup A_2} &\left|J_n(\mathbf{x},S_n+\frac{t}{\sqrt{n}})-\pi(S_n+\frac{t}{\sqrt{n}})\right| e^{L_n(S_n+\frac{t}{\sqrt{n}})-L_n(S_n)}dt \\ &\le \sup_{t\in A_1\cup A_2}\left|J_n(\mathbb{X},S_n+\frac{t}{\sqrt{n}})-\pi(S_n+\frac{t}{\sqrt{n}})\right| \int_{A_1\cup A_2} e^{L_n(S_n+\frac{t}{\sqrt{n}})-L_n(S_n)}dt. \end{aligned} \tag{A.8}$$

Recall that $\sqrt{n}(S_n - \theta_0) \xrightarrow{\mathcal{D}} N(0, 1/I(\theta_0))$, and therefore that

$$P_{\theta_0}\left(\{S_n + t/\sqrt{n};\ t \in A_1 \cup A_2\} \subset (\theta_0 - \delta_0, \theta_0 + \delta_0)\right) \to 1.$$

Moreover, it is established in the proof of Theorem 1.4.3 in Ghosh et al. (2003) that $\int_{\mathbb{A}_1} e^{L_n(S_n+s/\sqrt{n})-L_n(S_n)}\, ds = O_p(1)$. Therefore the right-hand-side of (A.8) goes to 0 in probability, by (A.5).

Turning our attention to A_3, notice that

$$\begin{aligned}
&\int_{A_3} \left| J_n(\mathbf{x}, S_n + \frac{t}{\sqrt{n}}) e^{L_n(S_n+\frac{t}{\sqrt{n}})-L_n(S_n)} - \pi(\theta_0) e^{-\frac{tI(\theta_0)}{2}} \right| dt \\
&\quad \le \int_{A_3} J_n(\mathbb{X}, S_n + \frac{s}{\sqrt{n}}) e^{L_n(S_n+\frac{s}{\sqrt{n}})-L_n(S_n)} + \int_{A_3} \pi(\theta_0) e^{-\frac{tI(\theta_0)}{2}} dt.
\end{aligned}$$

The second integral clearly goes to zero. The first integral is

$$\begin{aligned}
&\int_{A_3} J_n(\mathbb{X}, S_n + \frac{s}{\sqrt{n}}) e^{L_n(S_n+\frac{s}{\sqrt{n}})-L_n(S_n)} \\
&= \frac{1}{n} \sum_{j=1}^{n} \int_{A_3} J(X_j, S_n + \frac{s}{\sqrt{n}}) f(X_j, S_n + \frac{s}{\sqrt{n}}) e^{L_n(S_n+\frac{s}{\sqrt{n}})-L_n(S_n)-\log f(X_j, S_n+\frac{s}{\sqrt{n}})}\, ds
\end{aligned}$$

Notice that the definition of J implies $\int_{\mathbb{R}} J(X_j, S_n + s/\sqrt{n}) f(X_j, S_n + s/\sqrt{n}) = 1$. The convergence to zero therefore follows by the regularity assumptions (v) and (v'). Thus (A.6) is established. Finally, (5.3) is derived from (A.6) by straightforward algebra as exhibited in the proof of Theorem 1.4.3 of Ghosh et al. (2003).

References

Abdel-Karim, A. (2005). Applications of generalized inference. Ph.D. thesis, Colorado State University, Fort Collins, CO.

Allison, D. B., Gadbury, G. L., Heo, M., Fernández, J. R., Lee, C.-K., Prolla, T. A. and Weindruch, R. (2002). A mixture model approach for the analysis of microarray gene expression data. *Comput. Statist. Data Anal.* **39**, 1-20.

Barnard, G. A. (1995). Pivotal models and the fiducial argument. *Internat. Statist. Rev.* **63**, 309-323.

Behrens, W. V. (1929). Ein Beitrag zur Fehlerberchnung bei Wenigen Beobachtungen. *Landw. Jb.* **LXVIII**, 807-837.

Billingsley, P. (1995). *Probability and Measure.* John Wiley & Sons, New York, third edn.

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

Brown, L. D., Cai, T. T. and DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Statist.* **30**, 160-201.

Burch, B. D. and Iyer, H. K. (1997). Exact confidence intervals for a variance ratio (or heritability) in a mixed linear model. *Biometrics* **53**, 1318-1333.

Burdick, R. K., Borror, C. M. and Montgomery, D. C. (2005a). *Design and Analysis of Gauge R&R Studies*. ASA-SIAM Series on Statistics and Applied Probability. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Burdick, R. K., Park, Y.-J., Montgomery, D. C. and Borror, C. M. (2005b). Confidence intervals for misclassification rates in a gauge R&R study. *J. Quality Tech.* **37**, 294-303.

Cai, T. T. (2005). One-sided confidence intervals in discrete distributions. *J. Statist. Plann. Inference* **131**, 63-88.

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

Daniels, L., Burdick, R. K. and Quiroz, J. (2005). Confidence Intervals in a Gauge R&R Study with Fixed Operators. *J. Quality Tech.* **37**, 179-185.

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

Dawid, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. Ser. B* **35**, 189-233. With discussion by D. J. Bartholomew, A. D. McLaren, D. V. Lindley, Bradley Efron, J. Dickey, G. N. Wilkinson, A. P.Dempster, D. V. Hinkley, M. R. Novick, Seymour Geisser, D. A. S. Fraser and A. Zellner, and a reply by A. P. Dawid, M. Stone, and J. V. Zidek.

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

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

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

E, L., Hannig, J. and Iyer, H. K. (2008). Fiducial intervals for variance components in an unbalanced two-component normal mixed linear model. *J. Amer. Statist. Assoc.* **103**, 854-865.

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

Fisher, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society* **xxvi**, 528-535.

Fisher, R. A. (1933). The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proceedings of the Royal Society of London A* **139**, 343-348.

Fisher, R. A. (1935a). The fiducial argument in statistical inference. *Ann. Eugenics* **VI**, 91-98.

Fisher, R. A. (1935b). The logic of inductive inference. *J. Roy. Statist. Soc. B* **98**, 29-82.

Fraser, D. A. S. (1961). On fiducial inference. *Ann. Math. Statist.* **32**, 661-676.

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

Fraser, D. A. S. (1968). *The Structure of Inference*. John Wiley & Sons, New York-London-Sydney.

Fraser, D. A. S. (2006). Fiducial inference. In *The New Palgrave Dictionary of Economics* (Edited by S. Durlauf and L. Blume). Palgrave Macmillan, 2nd edition.

Ghosh, J. K. (1994). *Higher Order Assymptotics.* NSF-CBMS Regional Conference Series. Hayward: Institute of Mathematical Statistics.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics.* Springer Series in Statistics. Springer-Verlag, New York.

Glagovskiy, Y. S. (2006). *Construction of Fiducial Confidence Intervals For the Mixture of Cauchy and Normal Distributions.* Master's thesis, Department of Statistics, Colorado State University.

Grundy, P. M. (1956). Fiducial distributions and prior distributions: an example in which the former cannot be associated with the latter. *J. Roy. Statist. Soc. Ser. B* **18**, 217-221.

GUM (1995). *Guide to the Expression of Uncertainty in Measurement.* International Organization for Standardization (ISO), Geneva, Switzerland.

Hamada, M. and Weerahandi, S. (2000). Measurement system assessment via generalized inference. *J. Quality Tech.* **32**, 241-253.

Hannig, J. (1996). *On conditional distributions as limits of martingales.* Mgr. thesis, (in czech), Charles University, Prague, Czech Republic.

Hannig, J., E, L., Abdel-Karim, A. and Iyer, H. K. (2006a) Simultaneous fiducial generalized confidence intervals for ratios of means of lognormal distributions. *Austral. J. Statist.* **35**, 261-269.

Hannig, J., Iyer, H. K. and Patterson, P. (2006b) Fiducial generalized confidence intervals. *J. Amer. Statist. Assoc.* **101**, 254-269.

Hannig, J. and Lee, T. C. M. (2007). Generalized fiducial inference for wavelet regression. Tech. rep., Colorado State University.

Iyer, H. K. and Patterson, P. (2002). A recipe for constructing generalized pivotal quantities and generalized confidence intervals. Tech. Rep. 2002/10, Department of Statistics, Colorado State University.

Iyer, H. K., Wang, C. M. J. and Mathew, T. (2004). Models and confidence intervals for true values in interlaboratory trials. *J. Amer. Statist. Assoc.* **99**, 1060-1071.

Jeffreys, H. (1940). Note on the Behrens-Fisher formula. *Ann. Eugenics* **10**, 48-51.

Jeffreys, H. (1961). *Theory of Probability.* Clarendon Press, Oxford, third edn.

Le Cam, L. and Yang, G. L. (2000). *Asymptotics in Statistics.* Springer Series in Statistics. New York: Springer-Verlag, second edn.

Liao, C. T. and Iyer, H. K. (2004). A tolerance interval for the normal distribution with several variance components. *Statist. Sinica* **14**, 217-229.

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

McNally, R. J., Iyer, H. K. and Mathew, T. (2003). Tests for individual and population bioequivalence based on generalized p-values. *Statistics in Medicine* **22**, 31-53.

Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the Theory of Statistics.* McGraw-Hill, third edn.

Pounds, S. and Morris, S. W. (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics* **19**, 123601242.

Salome, D. (1998). *Staristical Inference via Fiducial Methods.* Ph.D. thesis, University of Groningen.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components*. John Wiley & Sons, New York.

Stevens, W. L. (1950). Fiducial limits of the parameter of a discontinuous distribution. *Biometrika* **37**, 117-129.

Tsui, K.-W. and Weerahandi, S. (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *J. Amer. Statist. Assoc.* **84**, 602-607.

Wang, C. M. and Iyer, H. K. (2005). Propagation of uncertainties in measurements using generalized inference. *Metrologia* **42**, 145-153.

Wang, C. M. and Iyer, H. K. (2006a). A generalized confidence interval for a measurand in the presence of type-A and type-B uncertainties. *Measurement* **39**, 856–863.

Wang, C. M. and Iyer, H. K. (2006b). Uncertainty analysis for vector measurands using fiducial inference. *Metrologia* **43**, 486-494.

Weerahandi, S. (1993). Generalized confidence intervals. *J. Amer. Statist. Assoc.* **88**, 899-905.

Weerahandi, S. (2004). *Generalized Inference in Repeated Measures*. Wiley, Hoboken, NJ.

Wilkinson, G. N. (1977). On resolving the controversy in statistical inference. *J. Roy. Statist. Soc. Ser. B* **39**, 119-171. With discussion.

Yeo, I.-K. and Johnson, R. A. (2001). A uniform strong law of large numbers for U-statistics with application to transforming to near symmetry. *Statist. Probab. Lett.* **51**, 63-69.

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

Department of Statistics and Operations Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3260, U.S.A.

E-mail: hannig@unc.edu

(Received November 2006; accepted December 2007)