Computational Statistics and Data Analysis 71 (2014) 849–858

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Computational issues of generalized fiducial inference

Jan Hannig [a], Randy C.S. Lai [b], Thomas C.M. Lee [b,*]

[a] Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3260, USA

[b] Department of Statistics, University of California at Davis, 4118 Mathematical Sciences Building, One Shields Avenue, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 10 April 2012
Received in revised form 1 March 2013
Accepted 4 March 2013
Available online 14 March 2013

Keywords:
Confidence distributions
Dempster–Shafer calculus
Generalized confidence intervals
Generalized fiducial inference

ABSTRACT

Generalized fiducial inference is closely related to the Dempster–Shafer theory of belief functions. It is a general methodology for constructing a distribution on a (possibly vector-valued) model parameter without the use of any prior distribution. The resulting distribution is called the generalized fiducial distribution, which can be applied to form estimates and confidence intervals for the model parameter. Previous studies have shown that such estimates and confidence intervals possess excellent frequentist properties. Therefore it is useful and advantageous to be able to calculate the generalized fiducial distribution, or at least to be able to simulate a random sample of the model parameter from it. For a small class of problems this generalized fiducial distribution can be analytically derived, while for some other problems its exact form is unknown or hard to obtain. A new computational method for conducting generalized fiducial inference without knowing the exact closed form of the generalized fiducial distribution is proposed. It is shown that this computational method enjoys desirable theoretical and empirical properties. Consequently, with this proposed method the applicability of generalized fiducial inference is enhanced.

1. Introduction

Fisher (1930) introduced the idea of fiducial probability and fiducial inference as an attempt to overcome what he saw as a serious deficiency of the Bayesian approach to inference: the use of a prior distribution on model parameters even when no prior information is available. Fiducial inference created some controversy once Fisher's contemporaries realized that, unlike earlier simple applications involving a single parameter, fiducial inference often led to procedures that were not exact in the frequentist sense and did not possess other properties claimed by Fisher (Lindley, 1958; Zabell, 1992). An interested reader can consult Section 2 of Hannig (2009) for a discussion of the history of fiducial inference and a more complete list of references.

Tsui and Weerahandi (1989) and Weerahandi (1993) proposed a new approach for constructing hypothesis tests using the concept of generalized *P*-values and generalized confidence intervals. Hannig et al. (2006) established a direct connection between fiducial intervals and generalized confidence intervals and proved the asymptotic frequentist correctness of such intervals. These ideas were unified for parametric problems in Hannig (2009) without requiring any group structure related to the model. This unification is termed *generalized fiducial inference* and has been found to have excellent theoretical and empirical properties for a number of practical applications (E et al., 2008; Hannig and Lee, 2009; Wandler and Hannig, 2011).

* Corresponding author. Tel.: +1 530 554 1357; fax: +1 530 752 7099.
E-mail addresses: jan.hannig@unc.edu (J. Hannig), rcslai@ucdavis.edu (R.C.S. Lai), tcmlee@ucdavis.edu (T.C.M. Lee).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.03.003

The Dempster–Shafer theory of belief functions (Dempster, 2008; Shafer, 2011) is closely related to the generalized fiducial inference. In both cases a relationship between the data, $\mathbf{X}$, and the parameters, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is expressed in a functional form; i.e.,

$$\mathbf{X} = \mathbf{G}(\boldsymbol{\theta}, \mathbf{U}), \tag{1.1}$$

where $\mathbf{U}$ is the random component of the structural equation, a random variable or vector whose distribution is completely known and independent of any parameters. The Eq. (1.1) is typically called data-generating equation, structural equation or a-equation.

After observing a fixed realized value of $\mathbf{X}$, say $\mathbf{x}_0$, one can invert the Eq. (1.1) to define a belief function on the parameter space $\boldsymbol{\Theta}$. To explain the idea behind the formal definition of this belief function, suppose first that the structural relation (1.1) can be inverted, solved for θ, and the inverse $\mathbf{Q}(\mathbf{x}_0, \mathbf{u})$ always exists. That is, for any observed $\mathbf{x}_0$ and for all $\mathbf{u}$, there is always a unique $\boldsymbol{\theta}$ solving $\mathbf{x}_0 = \mathbf{G}(\boldsymbol{\theta}, \mathbf{u})$. (The more realistic case where the inverse $\mathbf{Q}$ does not exist will be discussed in the next section.) Since the distribution of $\mathbf{U}$ is completely known, one can always generate a random sample $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_M$ from it. This random sample of $\mathbf{U}$ is transformed into a random sample of $\boldsymbol{\theta}$ via the inverse $\mathbf{Q}$: $\{\tilde{\boldsymbol{\theta}}_1 = \mathbf{Q}(\mathbf{x}_0, \tilde{\mathbf{u}}_1), \ldots, \tilde{\boldsymbol{\theta}}_M = \mathbf{Q}(\mathbf{x}_0, \tilde{\mathbf{u}}_M)\}$, and the resulting random sample $\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_M$ can be used to obtain estimates and approximate confidence intervals for $\boldsymbol{\theta}$. From this one can see that a probability density function $r(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$ is implicitly defined. The corresponding distribution is termed the *generalized fiducial distribution*, which, in this case, is equivalent to the belief function.

This recipe defines a joint distribution on the parameter space. When making inference about individual parameters, we then use the marginal of this joint distribution. It is well known that marginalization together with different structural equations can lead to non-uniqueness; i.e., several different marginal fiducial distributions, (e.g., Dawid et al., 1973). We do not see this as a problem because our goal is not to define the unique fiducial distribution. Instead, we are aiming to define a distribution on the parameter space with good inferential properties. Approximate confidence intervals based on the marginal fiducial distribution lead often to asymptotically correct coverage (Hannig, 2009, 2013). More importantly, such confidence intervals have been shown to have very good small sample properties in a number of applied problems. For computational reasons, we also recommend using structural equations that have simple form so that the calculations described in Section 3 can be done in a closed form; see for example Section 4.3. More discussion on the choice of structural equation can be found in Section 5 of Hannig (2013).

We remark that there is a new exciting inferential approach developed by Zhang and Liu (2011) that in some situations does not require a computation of the whole distribution, making it attractive for various problems. However, this method seems to have a requirement that the structural equation is invertible almost surely (Zhang and Liu, 2011, Theorem 3.1). Such an assumption is not reasonable when the dimension of the minimal sufficient statistics is larger than the number of parameters; e.g., the Cauchy regression problem to be discussed below.

The density function $r(\boldsymbol{\theta})$ thus plays an important role in the generalized fiducial approach to data analysis: it can be applied to derive estimates and construct confidence intervals for the parameter $\boldsymbol{\theta}$, in a similar manner as the posterior density function in the Bayesian paradigm. However, for many statistical problems, the exact form of $r(\boldsymbol{\theta})$ cannot be easily calculated. A major contribution of this article is to propose practical methods for computing integrals of $r(\boldsymbol{\theta})$ when a closed form expression for $r(\boldsymbol{\theta})$ is not readily available. Consequently, this article greatly enhances the applicability of generalized fiducial inference for statistical problems.

The rest of this article is organized as follows. Firstly some background material is presented in Section 2. Then some computational ideas are developed in Section 3 for simulating a random sample from $r(\boldsymbol{\theta})$ when the exact form of $r(\boldsymbol{\theta})$ is unknown. Section 4 provides some simulation studies to demonstrate the advantages of the proposed computational approach. Lastly, concluding remarks are offered in Section 5.

2. Background

This section provides some essential background material. First recall that in the discussion above the inverse $\mathbf{Q}$ is assumed to exist. In practice this is likely to be wrong and the inverse $\mathbf{Q}$ does not exist. This can happen for two opposing reasons: for some value of $\mathbf{x}_0$ and $\mathbf{u}$, either there is more than one $\boldsymbol{\theta}$, or there is no $\boldsymbol{\theta}$ satisfying $\mathbf{x}_0 = \mathbf{G}(\boldsymbol{\theta}, \mathbf{u})$. The first situation can be dealt with by using the mechanics of Dempster–Shafer calculus (Dempster, 2008); see also Section 4 of Hannig (2009). The main idea is that the belief function for any set has three components summarizing (i) the strength evidence for the set, (ii) the strength of the evidence against the set, and (iii) the strength of the evidence that is inconclusive. Hannig (2013) shows that in many statistical problems of practical interest the portion of the evidence that is inconclusive is asymptotically negligible.

For the second situation where no $\boldsymbol{\theta}$ satisfies $\mathbf{x}_0 = \mathbf{G}(\boldsymbol{\theta}, \mathbf{u})$, Hannig (2009) suggests removing the values of $\mathbf{u}$ for which there is no solution from the sample space and then re-normalizing the probabilities; i.e., using the distribution of $\mathbf{U}$ conditional on the event that "there is at least one $\boldsymbol{\theta}$ solving the equation $\mathbf{x}_0 = \mathbf{G}(\boldsymbol{\theta}, \mathbf{U})$". The rationale for this choice is that we know that the observed data $\mathbf{x}_0$ were generated using some fixed unknown $\boldsymbol{\theta}_0$ and $\mathbf{u}_0$; i.e., $\mathbf{x}_0 = \mathbf{G}(\boldsymbol{\theta}_0, \mathbf{u}_0)$. The information that the solution of the equation $\mathbf{x}_0 = \mathbf{G}(\boldsymbol{\theta}, \mathbf{U})$ exists for the true $\mathbf{U} = \mathbf{u}_0$ is available to us in addition to knowing the distribution of $\mathbf{U}$. The values of $\mathbf{u}$ for which $\mathbf{x}_0 = \mathbf{G}(\cdot, \mathbf{u})$ does not have a solution could not be the true $\mathbf{u}_0$ hence only the values of $\mathbf{u}$ for which there is a solution should be considered in the definition of the generalized fiducial

distribution, which leads to the conditioning. This step is closely related to the Dempster's rule of recombination (Dempster, 2008). See Section 2 of Hannig (2013) for a mathematically precise definition.

While the recipe for defining a belief function is conceptually straightforward, computational implementation is usually complicated even for simple problems. Moreover, the set of $\mathbf{u}$ for which the solution exists has probability zero in many practical situations; e.g., most problems involving absolutely continuous random variables. Conditioning on such a set of probability zero will therefore lead to non-uniqueness due to Borel paradox (e.g., Casella and Berger, 2002, Section 4.9.3).

Hannig (2013) provides an attractive resolution of the Borel Paradox by limit of discretizations. In particular, write the data generating equation as $\mathbf{G} = (g_1, \ldots, g_n)$ so that $X_i = g_i(\boldsymbol{\theta}, \mathbf{U})$ for $i = 1, \ldots, n$. Assume that $\mathbf{U} = (U_1, \ldots, U_n)$ is i.i.d. Uniform(0,1) and the parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ is p-dimensional. Then under some differentiability assumptions, Hannig (2013) shows that the generalized fiducial distribution is absolutely continuous with density

$$r(\boldsymbol{\theta}) = \frac{J(\mathbf{x}, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} J(\mathbf{x}, \boldsymbol{\theta}') f(\mathbf{x}, \boldsymbol{\theta}')\, d\boldsymbol{\theta}'}, \tag{2.1}$$

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

$$J(\mathbf{x}, \boldsymbol{\theta}) = \sum_{\substack{\mathbf{i}=(i_1,\ldots,i_p) \\ 1 \le i_1 < \cdots < i_p \le n}} \left| \det \left(\left(\frac{\mathbf{d}}{\mathbf{dx}} \mathbf{G}^{-1}(\mathbf{x}, \boldsymbol{\theta}) \right)^{-1} \frac{\mathbf{d}}{\mathbf{d}\boldsymbol{\theta}} \mathbf{G}^{-1}(\mathbf{x}, \boldsymbol{\theta}) \right)_{\mathbf{i}} \right|. \tag{2.2}$$

Here the sum goes over all p-tuples of indexes $\mathbf{i} = (1 \le i_1 < \cdots < i_p \le n) \subset \{1, \ldots, n\}$, $\mathbf{d}\mathbf{G}^{-1}(\mathbf{x}, \boldsymbol{\theta})/\mathbf{d}\boldsymbol{\theta}$ and $\mathbf{d}\mathbf{G}^{-1}(\mathbf{x}, \boldsymbol{\theta})/\mathbf{dx}$ are the $n \times p$ and $n \times n$ Jacobian matrixes respectively, and for any $n \times p$ matrix, $(A)_{\mathbf{i}}$ is the $p \times p$ matrix comprised of the rows $i_1, \ldots, i_p$ of A.

If the observations are from an i.i.d. univariate absolutely continuous distribution there is a natural choice of the structural equation (1.1). Let $F(x, \boldsymbol{\theta})$ and $f(x, \boldsymbol{\theta})$ be the distribution and density functions respectively, and $F^{-1}(\boldsymbol{\theta}, u) = \inf_x\{F(x, \boldsymbol{\theta}) \ge u\}$ is defined as the usual pseudo-inverse. The natural structural equation in this case is

$$X_i = F^{-1}(\boldsymbol{\theta}, U_i), \quad i = 1, \ldots, n, \tag{2.3}$$

where U_i are i.i.d. Uniform(0,1). The inverse of the structural equation $\mathbf{u} = \mathbf{G}^{-1}(\mathbf{x}, \boldsymbol{\theta})$ is

$$u_i = F(x_i, \boldsymbol{\theta}), \quad i = 1, \ldots, n$$

and the numerator of generalized fiducial density in (2.1) simplifies to

$$J(\mathbf{x}, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta}) = \sum_{\substack{\mathbf{i}=(i_1,\ldots,i_p) \\ 1 \le i_1 < \cdots < i_p \le n}} \left| \det \begin{pmatrix} \nabla_{\boldsymbol{\theta}} F(x_{i_1}, \boldsymbol{\theta}) \\ \vdots \\ \nabla_{\boldsymbol{\theta}} F(x_{i_p}, \boldsymbol{\theta}) \end{pmatrix} \right| \prod_{j \notin \mathbf{i}} f(x_j, \boldsymbol{\theta}), \tag{2.4}$$

where $\nabla_{\boldsymbol{\theta}} F(x, \boldsymbol{\theta})$ is the gradient; i.e., the row vector of partial derivatives computed with respect to elements of $\boldsymbol{\theta}$. Notice that if $n \ge p = 1$ then (2.1) and (2.4) agree with the proposal of Dempster (1963).

When the Jacobians in (2.2) or (2.4) can be evaluated in closed form one can use standard tools such as direct numerical evaluation, Markov Chain Monte Carlo, importance sampling or sequential Monte Carlo when implementing the formula (2.1). Examples of such successful implementation are Hannig et al. (2006); E et al. (2008); Hannig (2009); Hannig and Lee (2009); Wandler and Hannig (2011); Cisewski and Hannig (2012).

When the Jacobian cannot be computed in a closed form, one can attempt to compute it numerically, e.g., Wang et al. (2012). However, such computations can become prohibitively expensive even for moderate p as each evaluation would require a computation a large number of determinants of $p \times p$ Jacobians. As a main contribution of this article, next section develops some computational ideas specifically designed for such a situation.

3. Computational formulas

Let us first consider the i.i.d. setup of (2.3). Denote the fixed observed data value by $\mathbf{x}$ and its order statistic by $(x_{(1)}, \ldots, x_{(n)})$. Using the notation of the previous section let us assume that for each set of indexes $\mathbf{i} = (i_1, \ldots, i_p)$, $1 \le i_1 < \cdots < i_p \le n$, the p-equations $F(x_i, \boldsymbol{\theta}) = u_i$, $i \in \mathbf{i}$ have unique solution $\boldsymbol{\theta}$ for all $\mathbf{u} = (u_1, \ldots, u_p) \in \mathcal{S}_{\mathbf{i}} \subset (0, 1)^p$, and no solution if $\mathbf{u} \notin \mathcal{S}_{\mathbf{i}}$, where $\mathcal{S}_{\mathbf{i}}$ have positive Lebesgue measure. Denote this solution by $Q_{\mathbf{i}}(\mathbf{u})$ and assume that it is continuous. When using formulas (2.1) and (2.4) in practice, one needs to approximate integrals

$$\begin{aligned} \int_A J(\mathbf{x}, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta})\, d\boldsymbol{\theta} &= \sum_{\mathbf{i}} \int_A \left| \det \left(\nabla_{\boldsymbol{\theta}} F(x_{i_1}, \boldsymbol{\theta}), \ldots, \nabla_{\boldsymbol{\theta}} F(x_{i_p}, \boldsymbol{\theta}) \right) \right| \prod_{j \notin \mathbf{i}} f(x_j, \boldsymbol{\theta})\, d\boldsymbol{\theta} \\ &= \sum_{\mathbf{i}} \int_{Q_{\mathbf{i}}^{-1}(A)} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}))\, d\mathbf{u}. \end{aligned} \tag{3.1}$$

We will now discuss several options on how to approximate such integrals in practice.

3.1. Intrusive method

If the set A is closed and simply connected, then by the mean value theorem there is a point $\boldsymbol{\theta}_{\mathbf{i}} \in A$ such that

$$\int_{Q_{\mathbf{i}}^{-1}(A)} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}))\, d\mathbf{u} = \text{volume}(Q_{\mathbf{i}}^{-1}(A)) \prod_{j \notin \mathbf{i}} f(x_j, \boldsymbol{\theta}_{\mathbf{i}}). \tag{3.2}$$

If in addition A is a simplex spanned by points $\{\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_p\}$ we can approximate (3.2) by

$$\begin{aligned}&\int_{Q_{\mathbf{i}}^{-1}(A)} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}))\, d\mathbf{u}\\ &\quad\approx (p!)^{-1} \left| \det \begin{pmatrix} F(x_{i_1}, \boldsymbol{\theta}_1) - F(x_{i_1}, \boldsymbol{\theta}_0) & \cdots & F(x_{i_p}, \boldsymbol{\theta}_1) - F(x_{i_p}, \boldsymbol{\theta}_0) \\ \vdots & \vdots & \vdots \\ F(x_{i_1}, \boldsymbol{\theta}_p) - F(x_{i_1}, \boldsymbol{\theta}_0) & \cdots & F(x_{i_p}, \boldsymbol{\theta}_p) - F(x_{i_p}, \boldsymbol{\theta}_0) \end{pmatrix} \right| \prod_{j \notin \mathbf{i}} f_{X_j}(x_j, \tilde{\boldsymbol{\theta}}).\end{aligned}$$

Here we approximated the volume$Q_{\mathbf{i}}^{-1}(A)$ by the volume of the simplex spanned by the transformations of the vertexes of A and replaced the point $\boldsymbol{\theta}_{\mathbf{i}}$ with the center point $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_0 + \cdots + \boldsymbol{\theta}_p)/p$. Such a numerical scheme would be exact for linear functions (e.g., Ciarlet, 1978, p. 182).

Additionally, we can estimate the sum over all $\mathbf{i}$ by an average as the constant $\binom{n}{p}^{-1}$ would be applied to both numerator and denominator of (2.1). Next we estimate the average by an average of a random sample of size k of p-tuples of indexes $\mathbf{i}_1, \ldots, \mathbf{i}_K$; here $\mathbf{i}_k = \{1 \le i_{k,1} < \cdots < i_{k,p} \le n\}$. Thus for a simplex A spanned by $\{\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_p\}$ we approximate

$$\begin{aligned}&\binom{n}{p}^{-1} \int_A J(\mathbf{x}, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta})\, d\boldsymbol{\theta}\\ &\quad\approx K^{-1}(p!)^{-1} \sum_{k=1}^{K} \left| \det \begin{pmatrix} F(x_{i_{k,1}}, \boldsymbol{\theta}_1) - F(x_{i_{k,1}}, \boldsymbol{\theta}_0) & \cdots & F(x_{i_{k,p}}, \boldsymbol{\theta}_1) - F(x_{i_{k,p}}, \boldsymbol{\theta}_0) \\ \vdots & \vdots & \vdots \\ F(x_{i_{k,1}}, \boldsymbol{\theta}_p) - F(x_{i_{k,1}}, \boldsymbol{\theta}_0) & \cdots & F(x_{i_{k,p}}, \boldsymbol{\theta}_p) - F(x_{i_{k,p}}, \boldsymbol{\theta}_0) \end{pmatrix} \right| \prod_{j \notin \mathbf{i}_k} f_{X_j}(x_j, \tilde{\boldsymbol{\theta}}).\end{aligned} \tag{3.3}$$

This approximation is unbiased. Moreover, its quality improves for large n as the summands in (2.4) are U-statistics based on $\mathbf{X}$ (Hannig, 2009), and consequently the left-hand-side of (3.3) converges as $n \to \infty$ to the expected value $E \int_A J(\mathbf{x}, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta}$ a.s. Similarly the right-hand-side of (3.3) converges to the same limit as $K \to \infty$ and $n \to \infty$.

The approximation (3.3) is most useful when the parameter space is partitioned into simplexes. Such partitions are available in the numerical mathematics literature (e.g., Brandts et al., 2009). Once we have the simplex partition of the parameter space we can then implement a Metropolis–Hastings chain on the simplexes. In particular, if the simplex A spanned by $\{\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_p\}$ is the current state of the Metropolis–Hastings chain we propose the next state as one of the $p+1$ simplexes that share a face with A. Denote this proposed simplex by A' and assume it is spanned by $\{\boldsymbol{\theta}'_0, \ldots, \boldsymbol{\theta}'_p\}$. The proposal is then accepted/rejected based on the usual Metropolis–Hastings ratio computed from (2.1) and (3.3); i.e., it is accepted with probability

$$\frac{\sum_{k=1}^{K} V_k(\boldsymbol{\theta}'_0, \ldots, \boldsymbol{\theta}'_p) \prod_{j \notin \mathbf{i}_k} f_{X_j}(x_j, \tilde{\boldsymbol{\theta}}')}{\sum_{k=1}^{K} V_k(\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_p) \prod_{j \notin \mathbf{i}_k} f_{X_j}(x_j, \tilde{\boldsymbol{\theta}})},$$

where $\tilde{\boldsymbol{\theta}}' = (\boldsymbol{\theta}'_0 + \cdots \boldsymbol{\theta}'_p)/p$ and

$$V_k(\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_p) = \left| \det \begin{pmatrix} F(x_{i_{k,1}}, \boldsymbol{\theta}_1) - F(x_{i_{k,1}}, \boldsymbol{\theta}_0) & \cdots & F(x_{i_{k,p}}, \boldsymbol{\theta}_1) - F(x_{i_{k,p}}, \boldsymbol{\theta}_0) \\ \vdots & \vdots & \vdots \\ F(x_{i_{k,1}}, \boldsymbol{\theta}_p) - F(x_{i_{k,1}}, \boldsymbol{\theta}_0) & \cdots & F(x_{i_{k,p}}, \boldsymbol{\theta}_p) - F(x_{i_{k,p}}, \boldsymbol{\theta}_0) \end{pmatrix} \right|.$$

This method lowers the number of determinants of Jacobian one needs to calculate by eliminating the need to compute a numerical approximation to a Jacobian at every location used by evaluating volumes of moderate size simplexes. However, the need to then employ Metropolis–Hastings like algorithms makes it somewhat less desirable. In the next subsection we discuss a method that avoids both computation of Jacobian like determinants and Monte Carlo computations.

3.2. Non-intrusive method

In this section we propose a method for evaluating the generalized fiducial distribution $r(\boldsymbol{\theta})$ without computing any determinants. To this end we use (3.1) to compute the denominator of (2.1). In particular (3.1) simplifies to

$$\int_{\Theta} J(\mathbf{x}, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta})\, d\boldsymbol{\theta} = \sum_{\mathbf{i}} \int_{\mathcal{S}_{\mathbf{i}}} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}))\, d\mathbf{u}. \tag{3.4}$$

See beginning of Section 3 for the definition of $\mathcal{S}_{\mathbf{i}}$.

To numerically compute the integrals on the right-hand-side of (3.4) consider a partition $E_1, \ldots, E_M$ of $(0, 1)^p$. To achieve accuracy the diameter of the elements E_i should be small enough which could potentially lead to a lengthy calculation. Fortunately our extensive numerical experience suggests that only a relatively small number of elements near the maximum of the likelihood function are needed in order to get a good approximation of the integration.

This suggests the following approach. For each $\mathbf{i}$ compute

$$Q_{\mathbf{i}}^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}) = (F(x_{i_1}, \hat{\boldsymbol{\theta}}_{\mathrm{MLE}}), \ldots, F(x_{i_p}, \hat{\boldsymbol{\theta}}_{\mathrm{MLE}})),$$

where $\hat{\boldsymbol{\theta}}_{\mathrm{MLE}}$ is the maximum likelihood estimator (MLE) or a similar estimator for $\boldsymbol{\theta}$. For elements $E_l \subset [0, 1]^p$ near $Q_{\mathbf{i}}^{-1}(\hat{\boldsymbol{\theta}}_{\mathrm{MLE}})$ compute a numerical approximation to

$$\int_{E_l \cap \mathcal{S}_{\mathbf{i}}} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}))\, d\mathbf{u}.$$

With $\tilde{\mathbf{u}}_l$ the center of E_l, one such numerical approximation is $\mathrm{volume}(E_l) \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\tilde{\mathbf{u}}_l))$, provided $\tilde{\mathbf{u}}_l \in \mathcal{S}_{\mathbf{i}}$ and 0 otherwise. This requires computing the inverse $Q_{\mathbf{i}}(\tilde{\mathbf{u}}_l)$. Notice that due to continuity the already computed values of $Q_{\mathbf{i}}(\tilde{\mathbf{u}}_k)$ at a neighboring element can be used as the starting point for the numerical solver to speed up the computations. In the rest of this manuscript we will use the notation $f(x_j, Q_{\mathbf{i}}(\tilde{\mathbf{u}}_k)) = 0$ if the inverse $Q_{\mathbf{i}}(\tilde{\mathbf{u}}_k)$ does not exist.

In order to compute $\int_A J(\mathbf{x}, \boldsymbol{\theta}) f(\mathbf{x}, \boldsymbol{\theta})\, d\boldsymbol{\theta}$, consider for each $\mathbf{i}$

$$\begin{aligned} \sum_{E_l \subset Q_{\mathbf{i}}^{-1}(A)} \int_{E_l} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}))\, d\mathbf{u} &\le \int_{Q_{\mathbf{i}}^{-1}(A)} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}))\, d\mathbf{u} \\ &\le \sum_{E_l \cap Q_{\mathbf{i}}^{-1}(A) \neq \emptyset} \int_{E_l} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}))\, d\mathbf{u}. \end{aligned} \tag{3.5}$$

When using (3.5) recall that we have already calculated the inversion $Q_{\mathbf{i}}(\tilde{\mathbf{u}}_l)$ for all important elements E_l. We can then define the collection of elements $\mathcal{E}_{\mathbf{i}}(A) = \{E_l : Q_{\mathbf{i}}(\tilde{\mathbf{u}}_l) \in A\}$. Thus overall the generalized fiducial probability is approximated by

$$\int_A r(\boldsymbol{\theta}) d\boldsymbol{\theta} \approx \frac{\sum_{k=1}^{K} \sum_{E_l \in \mathcal{E}_{\mathbf{i}_k}(A)} \int_{E_l} \prod_{j \notin \mathbf{i}_k} f(x_j, Q_{\mathbf{i}_k}(\mathbf{u}))\, d\mathbf{u}}{\sum_{k=1}^{K} \sum_{l=1}^{M} \int_{E_l} \prod_{j \notin \mathbf{i}_k} f(x_j, Q_{\mathbf{i}_k}(\mathbf{u}))\, d\mathbf{u}}, \tag{3.6}$$

where $\mathbf{i}_1, \ldots, \mathbf{i}_K$ is random sample of size k of p-tuples of indexes and the integrals over the finite elements are approximated using numerical methods.

Proposition 3.1. *Assume that $F(x, u)$ is continuously differentiable in both variables. Then for each $\mathbf{i}$*

$$\sum_{E_l \in \mathcal{E}_{\mathbf{i}_k}(A)} \mathrm{volume}(E_l) \cdot \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\tilde{\mathbf{u}})) \to \int_{\mathcal{S}_{\mathbf{i}} \cap A} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}))\, d\mathbf{u}$$

as $\max_l \mathrm{diam} E_l \to 0$.

The proof of this proposition follows from the well-known results for numerical integration (Davis and Rabinowitz, 1984).

As a consequence, the approximation will for each k converge to the integral it is targeting as $\max_i \mathrm{diam} E_i \to 0$. Moreover, strong law of large numbers then provides convergence of the right hand side of (3.6) to the left hand side of (3.6) as $k \to \infty$.

The exact speed of convergence depends on the choice of the partition. The natural regular partition $E_1, \ldots, E_N$ consisting of cubes of size $1/N^{1/d}$ converges only at the rate $N^{1/d}$ (Davis and Rabinowitz, 1984, Section 5.5). It is often recommended to improve the speed of convergence by the use of so-called low discrepancy sequences such as the Hammersley set. In particular for smooth enough densities if $\mathbf{u}_k$ are members of the Hammersley sequence,

$$N^{-1} \sum_{k=1}^{N} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}_k)) \to \int_{\mathcal{S}_{\mathbf{i}}} \prod_{j \notin \mathbf{i}} f(x_j, Q_{\mathbf{i}}(\mathbf{u}))\, d\mathbf{u}$$

at the rate of $(\log N)^{d-1} N^{-1}$.

3.3. Algorithm for the non-intrusive method

This subsection summarizes the major steps of the above non-intrusive method for approximating the generalized fiducial probability $\int_A r(\boldsymbol{\theta})d\boldsymbol{\theta}$ in the form of step-by-step instructions.

First obtain $\hat{\boldsymbol{\theta}}_{\text{MLE}}$, and sample K sets of p-tuples of indexes $\mathbf{i}_k = \{1 \le i_{k,i} < \cdots < i_{k,p} \le n\}$, $k = 1, \ldots, K$. For each k, perform the following steps:

1. Compute $Q_{\mathbf{i}_k}^{-1}(\hat{\boldsymbol{\theta}}_{\text{MLE}}) = (F(x_{i_1}, \hat{\boldsymbol{\theta}}_{\text{MLE}}), \ldots, F(x_{i_p}, \hat{\boldsymbol{\theta}}_{\text{MLE}}))$.
2. Partition $(0, 1)^p$ into hyper-cubes (e.g., squares for 2-d, cubes for 3-d) $E_{k,l}$, $1 \le l \le M$. The sizes of these hyper-cubes should be small enough to ensure numerical accuracy; hyper-cubes near $Q_{\mathbf{i}_k}^{-1}(\hat{\boldsymbol{\theta}}_{\text{MLE}})$ can be made smaller than the hypercubes further away to gain speed. Denote the center of $E_{k,l}$ as $\tilde{\mathbf{u}}_{k,l}$.
3. Compute $\tilde{\boldsymbol{\theta}}_{k,l} = Q_{\mathbf{i}_k}(\tilde{\mathbf{u}}_{k,l})$; i.e., the solution to the systems of equations
$$\left(F(x_{i_{k,1}}, \tilde{\boldsymbol{\theta}}_{k,l}), \ldots, F(x_{i_{k,p}}, \tilde{\boldsymbol{\theta}}_{k,l})\right)' = \tilde{\mathbf{u}}_{k,l}.$$
Numerical methods such as Newton's method can be employed here. We note that as solution may not exist for some $\mathbf{u}$, we only restrict ourselves to those $E_{k,l}$'s that provide solutions to the equations.
4. Calculate $J_{k,l} = \text{Volume}(E_{k,l}) \prod_{j \notin \mathbf{i}_k} f(x_j, \tilde{\boldsymbol{\theta}}_{k,l})$.

The generalized fiducial probability $\int_A r(\boldsymbol{\theta})d\boldsymbol{\theta}$ is then approximated by

$$\frac{\sum_{k=1}^{K}\sum_{l=1}^{M} J_{k,l}\mathbf{1}(\tilde{\boldsymbol{\theta}}_{k,l} \in A)}{\sum_{k=1}^{K}\sum_{l=1}^{M} J_{k,l}},$$

and from which a random sample of $\boldsymbol{\theta}$ can be (approximately) simulated to form point estimates and confidence intervals for $\boldsymbol{\theta}$.

The algorithm can be modified accordingly if we choose to use a low discrepancy sequence for integration.

4. Simulation study

To demonstrate the feasibility of the non-intrusive method for approximating the generalized fiducial distribution we will demonstrate it on three different problems: estimation for the two parameter Gamma distribution, estimation for the three parameter Weibull distribution, and linear regression with Cauchy errors. In each of these cases we performed a simulation study in which we compared the repeating sampling frequentist performance of the generalized fiducial distribution implemented using the non-intrusive method with other methods used in the literature for these problems.

The non-intrusive methods works with the joint generalized fiducial distribution. In order to find a confidence interval for a single marginal parameters, appropriate rectangular sets A are used, e.g. $A = (-\infty, a) \times \mathbb{R}$. In other words, in order to construct a confidence interval of a parameter, we first sample from the joint fiducial distribution then marginalize the samples. The simulation results show generalized fiducial method's favorable performance in terms of coverage and length of approximate confidence intervals.

Throughout the simulation section we compare our method with the MLE and/or MPS (Maximum Product of Spacing) methods to be described below. These two methods construct confidence intervals based on asymptotic distributions of likelihood or likelihood-like quantity. If other nuisance parameters are present, they are replaced by their consistent estimators.

Here we do not include any simulation results for the intrusive method, as in the examples considered the intrusive method is computationally slower then the non-intrusive method. It is because the intrusive method involves partitioning the space into a large number of simplexes and a computationally expensive MCMC procedure is required to visit such simplexes. Also, this MCMC based intrusive method that cannot be easily parallelized, while the non-intrusive method can be easily parallelized. This makes the non-intrusive method more suited for modern computer environments. Therefore in general the non-intrusive method is preferable, especially when there is a relatively fast way to compute the inverse $Q(\tilde{\mathbf{u}})$.

4.1. Gamma distribution

In this subsection we considered samples of size $n = \{3, 10, 30\}$ from a Gamma distribution with shape parameter $\alpha = \{0.5, 1, 5\}$ and rate parameter $\lambda = 1$. Because of the form of the Gamma distribution there is no closed form solution for either the maximum likelihood estimator nor the generalized fiducial density. Therefore this setting is a prime example where our algorithms will be useful. A variant of the intrusive method has been implemented in Wang et al. (2012) in the context of prediction. Here we report the results of the non-intrusive method.

Table 1
Empirical coverages, in %, of various 95% confidence intervals calculated from 500 independent data sets for the Gamma$(\alpha, \lambda) = (1, 1)$ simulation. Numbers in parentheses are median lengths of these confidence intervals.

n	α		λ	
	Fiducial	MLE	Fiducial	MLE
3	95 (3.31)	99 (4.22)	93 (4.58)	98 (5.96)
10	96 (1.60)	98 (1.78)	93 (1.88)	98 (2.28)
30	95 (0.92)	96 (0.95)	93 (1.14)	95 (1.22)

Table 2
Similar to Table 1 but for the Gamma$(\alpha, \lambda) = (0.5, 1)$ simulation.

n	α		λ	
	Fiducial	MLE	Fiducial	MLE
3	94 (2.08)	97 (2.25)	93 (7.38)	97 (7.54)
10	93 (0.76)	96 (0.82)	95 (2.39)	96 (2.52)
30	94 (0.41)	94 (0.43)	93 (1.25)	96 (1.36)

Table 3
Similar to Table 1 but for the Gamma$(\alpha, \lambda) = (5, 1)$ simulation.

n	α		λ	
	Fiducial	MLE	Fiducial	MLE
3	92 (19.29)	97 (26.19)	91 (3.89)	98 (5.19)
10	94 (8.95)	97 (10.39)	95 (1.84)	98 (2.15)
30	95 (5.00)	96 (5.21)	95 (1.05)	96 (1.10)

For each combination of sample size and parameter values, 500 independent data sets were generated. For each of these generated data sets, two approximate 95% confidence intervals were constructed for each of the parameters: one was based on the generalized fiducial distribution computed with the non-intrusive algorithm using $K = 100$ p-tuples and $M = 1000$ hypercubes. The other was based on the marginal asymptotic distribution of the MLE. Tables 1–3 list the empirical coverages and median lengths of these approximate 95% confidence intervals, based on the 500 independent data sets. These tables show that the fiducial based intervals maintain the stated coverage reasonably well while the MLE based intervals tend to be slightly conservative. Overall the fiducial based intervals are shorter.

We note that our generalized fiducial method is computationally more demanding than the classical MLE methods. However, it can be seen that the generalized fiducial method outperforms the classical MLE method when the sample size is relatively small. It suggests that the extra expense of computation is worthing for a better performance.

4.2. Three parameter Weibull distribution

In this subsection we considered samples of size $n = \{4, 10, 30\}$ from the three parameter Weibull distribution:

$$F(x, \alpha, \lambda, \theta) = (1 - e^{-\lambda^{\alpha}(x-\theta)^{\alpha}})I_{(\theta,\infty)}(x), \quad \lambda > 0, \alpha > 0.$$

In our simulation we have considered $\lambda = 1$, $\theta = 0$ and $\alpha = \{0.5, 1, 5\}$.

If the location parameter is fixed and known $\theta = 0$ the resulting two parameter Weibull distribution is a transformation of location scale family and therefore the generalized fiducial distribution is equivalent to a Bayesian posterior with reference prior (Berger et al., 2009), and is known to lead to exact frequentist confidence intervals.

The three parameter situation is much more interesting. The usual reference prior recipe leads to priors that give improper posteriors. The generalized fiducial distribution is always proper but has a very heavy tail in the θ parameter; the fiducial distribution of θ does not have expected value. This makes it very challenging to use standard MCMC techniques to sample from the generalized fiducial distribution, even though the closed form of the generalized fiducial density could be potentially computed. The non-intrusive approach allows us to alleviate some of these problems by a natural transformation of the problem to a problem of integrating on $(0, 1)^3$.

It is well-known that the maximum likelihood estimators could fail in estimating parameters in three-parameter Weibull distribution (Rockette et al., 1974; Smith, 1985; Nagatsuka et al., 2013) and could not therefore be used for comparison purposes. Three cases may occur:

1. $\alpha < 1$: the maximized likelihood is infinity and no local maximizer can ever be a consistent estimator;
2. $1 \leq \alpha < 2$: there is a positive probability that the density has no local maximum and the asymptotic distribution of the estimator does not follow any normal distribution; and
3. $\alpha \geq 2$: the weak regularity conditions are satisfied and the maximum likelihood estimators are asymptotic consistent, efficient and normal.

Table 4
Similar to Table 1 but for the Weibull$(\alpha, \lambda, \theta) = (1, 1, 0)$ simulation.

n	α		λ		θ	
	Fiducial	MPS	Fiducial	MPS	Fiducial	MPS
4	98 (10.55)	73 (2.60)	100 (26.96)	69 (5.75)	98 (6.94)	69 (2.15)
10	96 (2.46)	94 (1.70)	96 (2.21)	89 (3.15)	97 (1.50)	92 (0.97)
30	93 (0.70)	99 (1.14)	96 (0.91)	99 (1.81)	97 (0.20)	99 (0.42)

Table 5
Similar to Table 1 but for the Weibull$(\alpha, \lambda, \theta) = (0.5, 1, 0)$ simulation.

n	α		λ		θ	
	Fiducial	MPS	Fiducial	MPS	Fiducial	MPS
4	97 (3.61)	90 (1.34)	99 (50.30)	84 (31.09)	98 (3.77)	81 (2.38)
10	95 (0.59)	99 (0.82)	96 (4.05)	100 (10.56)	97 (0.18)	99 (0.52)
30	93 (0.29)	99 (0.58)	95 (1.71)	100 (4.90)	95 (0.013)	99 (0.088)

Table 6
Similar to Table 1 but for the Weibull$(\alpha, \lambda, \theta) = (5, 1, 0)$ simulation.

n	α		λ		θ	
	Fiducial	MPS	Fiducial	MPS	Fiducial	MPS
4	93 (27.19)	61 (61.47)	100 (27.53)	45 (0.056)	91 (4.26)	18 (0.69)
10	96 (42.81)	40 (8.70)	96 (3.20)	40 (0.51)	96 (7.84)	20 (0.48)
30	97 (40.70)	58 (6.25)	97 (1.80)	29 (0.30)	97 (7.63)	25 (0.34)

To avoid inconsistent estimators when $\alpha < 1$, it was proposed by Cheng and Amin (1983) to replace the likelihood by a measure of spacings between the observations. The resulting estimator is called the Maximum Product of Spacing (MPS) estimator. It is shown that the MPS estimator exists under more general situations and it is asymptotically equivalent to MLE if MLE exists. MPS has been successfully to construct confidence intervals and goodness-of-fit test (Cheng and Stephens, 1989; Cheng and Traylor, 1995).

For comparison purposes, both generalized fiducial inference with $K = 100$ and $M = 10^7$ and MPS were used to construct approximate confidence intervals in our simulations. Tables 4–6 report the empirical coverages and median lengths of the various confidence intervals. These tables show that the fiducial based confidence intervals are to be preferred, especially for small values of n.

The numerical issue cause by the location parameter θ led us to use a rather large M which in turn made this the slowest of all our examples. To further gain efficiency we implemented the Hammersley sequence computation described at the end of Section 3.2. A simulation study for one of the parameters setting took about a day on the University of North Carolina cluster computer. Though the computational expense for the generalized fiducial method is substantial, we still feel it is preferable to the rather erratic MPS method.

4.3. Simple linear regression with Cauchy errors

In this last simulation we considered the simple linear regression model with Cauchy errors,

$$Y_j = \beta_0 + \beta_1 X_j + \sigma Z_j,$$

where X_j is a covariate and Z_j are i.i.d. standard Cauchy; i.e., $F_z(z) = \frac{1}{2} + \frac{1}{\pi}\arctan(z)$.

This example shows that the non-intrusive algorithm can be simply extended to non i.i.d. data situations. For the current problem the function $Q_{\mathbf{i}}(\mathbf{u})$ for non-intrusive method is

$$Q_{\mathbf{i}}(\mathbf{u}) = (\mathbf{H}_{\mathbf{i}}'\mathbf{H}_{\mathbf{i}})^{-1}\mathbf{H}_{\mathbf{i}}'\mathbf{Y}_{\mathbf{i}},$$

where $\mathbf{X_i} = (X_{i_1}, \ldots, X_{i_p})'$, $\mathbf{Y_i} = (Y_{i_1}, \ldots, Y_{i_p})'$ and $\mathbf{H_i} = \left[1, \mathbf{X_i}, F_z^{-1}(\mathbf{u})\right]$. Similarly as in the case of i.i.d. data only a subset of $\mathbf{u} \in (0, 1)^3$ has a defined inverse Q. Indeed, if $Q_{\mathbf{i}}(F_z(\mathbf{z})) = (\beta_0, \beta_1, \sigma)'$, then $Q_{\mathbf{i}}(F_z(-\mathbf{z})) = (\beta_0, \beta_1, -\sigma)'$. Thus the set $\mathcal{S}$ of $\mathbf{u} = F_z(\mathbf{z})$ for which there is an inverse has measure $1/2$.

We considered samples of size $n = \{4, 10, 30\}$ with the design points X_j being generated from the standard normal distribution, $\beta_0 = 0$, $\beta_1 = \{1, 2, 5\}$ and $\sigma = 1$. In Tables 7–9 we report the performance of the generalized fiducial inference confidence intervals using $K = 100$ and $M = 10^4$, and the confidence intervals based on the asymptotics of MPS and MLE. The generalized fiducial based intervals show steady reliable performance while the competing methods are quite erratic.

A simulation for one of the parameter settings took about 6 to 7 h using a regular PC with an Intel(R) Core(TM) i7 3.33 GHz CPU (about 45s per dataset). The speed of the non-intrusive method, though slower than MLE and MPS, is still very reasonable due to the fact that the inverse $Q_{\mathbf{i}}(\tilde{u})$ has a closed form in this problem.

Table 7
Similar to Table 1 but for the Cauchy regression $(\beta_0, \beta_1, \sigma) = (0, 1, 1)$ simulation.

n	β_0			β_1			σ		
	Fiducial	MPS	MLE	Fiducial	MPS	MLE	Fiducial	MPS	MLE
4	94 (8.68)	94 (6.26)	0 (0.00011)	94 (12.03)	96 (6.79)	0 (0.00012)	96 (7.61)	94 (4.57)	0 (8.2e−05)
10	95 (2.63)	94 (2.35)	74 (1.41)	94 (3.10)	89 (2.56)	74 (1.47)	94 (2.53)	98 (2.22)	69 (1.33)
30	96 (1.14)	96 (1.23)	90 (0.96)	94 (1.20)	56 (1.26)	88 (0.97)	95 (1.09)	96 (1.20)	89 (0.94)

Table 8
Similar to Table 1 but for the Cauchy regression $(\beta_0, \beta_1, \sigma) = (0, 2, 1)$ simulation.

n	β_0			β_1			σ		
	Fiducial	MPS	MLE	Fiducial	MPS	MLE	Fiducial	MPS	MLE
4	94 (9.33)	94 (6.88)	0 (1e−04)	95 (12.17)	93 (7.77)	0 (0.00012)	96 (7.50)	99 (4.84)	0 (7.5e−05)
10	95 (2.63)	96 (2.83)	71 (1.35)	95 (3.05)	76 (2.99)	71 (1.41)	95 (2.42)	99 (2.68)	71 (1.26)
30	94 (1.13)	91 (1.40)	87 (0.92)	96 (1.20)	34 (1.42)	86 (0.93)	94 (1.12)	89 (1.38)	85 (0.90)

Table 9
Similar to Table 1 but for the Cauchy regression $(\beta_0, \beta_1, \sigma) = (0, 5, 1)$ simulation.

n	β_0			β_1			σ		
	Fiducial	MPS	MLE	Fiducial	MPS	MLE	Fiducial	MPS	MLE
4	95 (8.92)	91 (12.77)	0 (1e−04)	95 (11.76)	82 (14.53)	0 (0.00011)	94 (7.62)	99 (8.50)	0 (7.7e−05)
10	96 (2.53)	88 (5.50)	75 (1.30)	95 (2,93)	18 (5.80)	71 (1.37)	94 (2.34)	98 (5.24)	71 (1.22)
30	94 (1.13)	88 (3.00)	91 (0.96)	94 (1.20)	0.2 (3.09)	89 (0.95)	94 (1.10)	12 (2.94)	88 (0.93)

4.4. Discussion of simulation results

In the above numerical experiments the experimental settings (e.g., the distribution function and the corresponding parameter values) were carefully chosen with the hope to represent a large class of practical scenarios. However, just the same as any other simulation studies, the above numerical study could not cover all the possibilities that one may encounter in practice, and therefore caution must be exercised when empirical conclusions are being drawn from the results. Despite this, we believe that the following conclusions are reliable:

- The relative performances of the fiducial and classical (MLE and MPS) methods depend on the sample size and the shape of the underlying distribution.
- When the sample size is large, the fiducial and classical methods tend to give similar results. On the other hand, for small sized problems, the fiducial method could potentially outperform the classical methods by a large margin.
- The fiducial method also tends to provide better results when the underlying distributions are highly skewed or heavy-tailed.
- Very often the fiducial method is computationally more expensive than the classical methods. Therefore if the sample size is large and the underlying distribution is relatively symmetric without heavy-tails, one could use the classical methods, for the reason of time saving. Otherwise, one should use the fiducial method to achieve statistical accuracies.
- When possible we recommend the use of structural equations with simple closed form inverses, such as in the case of Cauchy regression, in order to speed up computations.

5. Conclusions

In this article we have proposed two practical methods for simulating the generalized fiducial samples: the intrusive method and the non-intrusive method. The intrusive method requires Metropolis–Hastings like algorithms to generate random samples from the target distribution, while the non-intrusive method avoids such Monte Carlo computations by transforming the generalized fiducial integral from the parameter space $\boldsymbol{\Theta}$ to $(0, 1)^p$. The non-intrusive method can be applied straightforwardly without computing the exact expression of the generalized fiducial distribution, nor any of its derivatives. Simulation results show that fiducial based confidence intervals produced by the non-intrusive method have very reliable frequentist properties. Quite often they outperform other classical methods by a large margin when the sample size is small. Since for many statistical problems the exact expression of the generalized fiducial distribution is unknown, the non-intrusive method greatly enhances the applicability of generalized fiducial inference.

Acknowledgments

The authors are most grateful to the reviewers for their useful comments, which led to a much improved version of the paper. The work of Hannig was partially supported by the National Science Foundation Grants 1007543 and 1016441. The work of Lee was partially supported by the National Science Foundation Grants 1007520, 1209226 and 1209232.

References

Berger, J.O., Bernardo, J.M., Sun, D., 2009. The formal definition of reference priors. The Annals of Statistics 37 (2), 905–938.

Brandts, J., Korotov, S., Křížek, M., Šolc, J., 2009. On nonobtuse simplicial partitions. SIAM Review 51 (2), 317–335.

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

Cheng, R., Amin, N., 1983. Estimating parameters in continuous univariate distributions with a shifted origin. Journal of the Royal Statistical Society: Series B 394–403.

Cheng, R., Stephens, M., 1989. A goodness-of-fit test using Moran's statistic with estimated parameters. Biometrika 76 (2), 385–392.

Cheng, R., Traylor, L., 1995. Non-regular maximum likelihood problems. Journal of the Royal Statistical Society: Series B 3–44.

Ciarlet, P.G., 1978. The Finite Element Method for Elliptic Problems. North-Holland Publishing Co., Amsterdam.

Cisewski, J., Hannig, J., 2012. Generalized fiducial inference for normal linear mixed models. The Annals of Statistics.

Davis, P.J., Rabinowitz, P., 1984. Methods of Numerical Integration, second ed. In: Computer Science and Applied Mathematics, Academic Press Inc., Orlando, FL.

Dawid, A.P., Stone, M., Zidek, J.V., 1973. Marginalization paradoxes in Bayesian and structural inference. Journal of the Royal Statistical Society. Series B. Methodological 35, 189–233.

Dempster, A.P., 1963. On direct probabilites. Journal of the Royal Statistical Society: Series B 25, 100–110.

Dempster, A.P., 2008. The Dempster–Shafer calculus for statisticians. International Journal of Approximate Reasoning 48, 365–377.

E, Lidong., Hannig, J., Iyer, H.K., 2008. Fiducial intervals for variance components in an unbalanced two-component normal mixed linear model. Journal of American Statistical Association 103 (482), 854–865.

Fisher, R.A., 1930. Inverse probability. Proceedings of the Cambridge Philosophical Society xxvi, 528–535.

Hannig, J., 2013. Generalized fiducial inference via discretization. Statistica Sinica 23, 489–514.

Hannig, J., 2009. On generalized fiducial inference. Statistica Sinica 19, 491–544.

Hannig, J., Iyer, H.K., Patterson, P., 2006. Fiducial generalized confidence intervals. Journal of American Statistical Association 101 (473), 254–269.

Hannig, J., Lee, T.C.M., 2009. Generalized fiducial inference for wavelet regression. Biometrika 96, 847–860.

Lindley, D.V., 1958. Fiducial distributions and Bayes' theorem. Journal of the Royal Statistical Society: Series B 20, 102–107.

Nagatsuka, H., Kamakura, T., Balakrishnan, N., 2013. A consistent method of estimation for the three-parameter Weibull distribution. Computational Statistics and Data Analysis 58, 210–226.

Rockette, H., Antle, C., Klimko, L., 1974. Maximum likelihood estimation with the Weibull model. Journal of the American Statistical Association 69, 246–249.

Shafer, G., 2011. A betting interpretation of Dempster–Shafer degrees of belief. International Journal of Approximate Reasoning 52, 127–136.

Smith, R., 1985. Maximum likelihood estimation in a class of nonregular cases. Biometrika 72 (1), 67–90.

Tsui, K.-W., Weerahandi, S., 1989. Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. Journal of the American Statistical Association 84 (406), 602–607.

Wandler, D.V., Hannig, J., 2011. Fiducial inference on the maximum mean of a multivariate normal distribution. Journal of Multivariate Analysis 102, 87–104.

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

Weerahandi, S., 1993. Generalized confidence intervals. Journal of the American Statistical Association 88 (423), 899–905.

Zabell, S.L., 1992. R. A. Fisher and the fiducial argument. Statistical Science 7 (3), 369–387.

Zhang, J., Liu, C., 2011. Dempster–Shafer inference with weak beliefs. Statistica Sinica 21 (2), 475–494.