Statistica Sinica **23** (2013), 489-514
doi:http://dx.doi.org/10.5705/ss.2011.102

GENERALIZED FIDUCIAL INFERENCE VIA DISCRETIZATION

Jan Hannig

The University of North Carolina at Chapel Hill

Abstract: In addition to the usual sources of error that have been long studied by statisticians, many data sets have been rounded off in some manner, either by the measuring device or storage on a computer. In this paper we investigate theoretical properties of generalized fiducial distribution introduced in Hannig (2009) for discretized data. Limit theorems are provided for both fixed sample size with increasing precision of the discretization, and increasing sample size with fixed precision of the discretization. The former provides an attractive definition of generalized fiducial distribution for certain types of exactly observed data overcoming a previous non-uniqueness due to Borel paradox. The latter establishes asymptotic correctness of generalized fiducial inference, in the frequentist, repeated sampling sense, for i.i.d. discretized data under very mild conditions.

Key words and phrases: Asymptotic properties, Bernstein-von Mises theorem, Dempster-Shafer calculus, generalized fiducial inference.

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. In the case of a one-parameter family of distributions, Fisher gave the following definition for a fiducial density $r(\theta)$ of the parameter based on a single observation x_0 for the case where the distribution function $F(x|\theta)$ is a function of θ decreasing from 1 to 0:

$$r(\theta) = -\frac{\partial F(x_0|\theta)}{\partial\theta}. \tag{1.1}$$

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)). More positively, Fraser (1968) developed a rigorous framework for making inferences along the lines of Fisher's fiducial inference assuming that the statistical model was coupled with an additional group structure, e.g., the location-scale model. Wilkinson (1977) argued that for complicated problems the fiducial distribution is not unique and should depend on the parameter of interest. Dawid and Stone (1982) provided further insight by studying certain situations where fiducial inference leads to exact confidence statements. Barnard (1995) proposed a view of fiducial distributions based on the pivotal approach that seems to eschew some of the problems reported in earlier literature. Dempster (2008) and Shafer (2011) discussed Dempster-Shafer calculus, which is closely related to fiducial inference. An interested reader can consult Section 2 of Hannig (2009) for a more thorough 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. These generalized confidence intervals have been found in many simulation studies to have good empirical frequentist properties, see Hannig, Iyer, and Patterson (2006) for references. Hannig, Iyer, and Patterson (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 good theoretical and empirical properties for a number of practical applications (E, Hannig, and Iyer (2008, 2009); Hannig and Lee (2009); Wandler and Hannig (2011, 2012a,b)).

Traditionally, the goal of fiducial inference was to formulate clear principles that would guide a statistician to a unique fiducial distribution. Generalized fiducial inference does not have such a goal. It treats the mechanics of generalized fiducial inference as a tool to define a distribution on the parameter space and uses this distribution to propose statistical procedures, e.g. approximate confidence intervals. The quality of the proposed procedures is then evaluated on their own merit using theoretical large sample properties and simulations.

Generalized fiducial inference begins with expressing the relationship between the data, $\mathbf{X}$, and the parameters, ξ, as

$$\mathbf{X} = G(\xi, \mathbf{U}), \tag{1.2}$$

where $G(\cdot, \cdot)$ is termed a structural equation, and $\mathbf{U}$ is the random component of the structural equation, a random variable or vector whose distribution is completely known and independent of any parameters. We intentionally leave the definition of the structural equation as general as possible. We offer some comments and suggestions on how to select a structural equation in Section 5.

A formal definition of a generalized fiducial distribution will be presented in Section 2. The purpose of the rest of this section is to give the reader a heuristic understanding of the ideas developed in this manuscript.

Let $\mathbf{x}_0$ be the fixed realized value of $\mathbf{X}$ generated using some fixed unobserved parameter ξ_0. To explain the idea behind the formal definition of generalized fiducial distribution, suppose first that the structural relation (1.2) can be inverted and the inverse $\mathbf{G}^{-1}(\mathbf{x}_0, \mathbf{u})$ always exists. That is, for observed $\mathbf{x}_0$ and all $\mathbf{u}$, there is the unique ξ solving $\mathbf{x}_0 = \mathbf{G}(\xi, \mathbf{u})$. As example of such a situation consider $\mathbf{x}_0$ a sample of size $n = 1$ from a location parameter family, $X = \xi + U$, or $\mathbf{x}_0$ a sample of size $n = 2$ generated using $X_1 = \xi_1 + \xi_2 U_1, X_2 = \xi_1 + \xi_2 U_2$ with $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2$. 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 $\{\tilde{\xi}_1 = G^{-1}(\mathbf{x}_0, \tilde{\mathbf{u}}_1), \ldots, \tilde{\xi}_M = G^{-1}(\mathbf{x}_0, \tilde{\mathbf{u}}_M)\}$, which is called the fiducial sample. The fiducial sample $\tilde{\xi}_1, \ldots, \tilde{\xi}_M$ is a sample from the fiducial distribution and can be used to obtain estimates and approximate confidence intervals for ξ.

The inverse $G^{-1}(\cdot\,,\cdot\,)$ often does not exist. This can happen under two situations: for some value of $\mathbf{x}_0$ and $\mathbf{u}$, either there is more than one ξ, or there is no ξ satisfying $\mathbf{x}_0 = \mathbf{G}(\xi, \mathbf{u})$.

The first situation can be dealt with by using the mechanics of Dempster-Shafer calculus (Dempster (2008)), see Section 4 of Hannig (2009) for more detail. A more practical solution is to select one of the several solutions using some possibly random mechanism. We provide some guidance on how this selection can be made in Section 2. see also Section 6 of Hannig (2009). In any case, we show in this paper that in many problems of practical interest the method of selection has only a second order effect on statistical inference.

For the second situation, 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 the "there is at least one ξ solving the equation $\mathbf{x}_0 = G(\xi, \mathbf{U})$". The rationale for this choice is that we know that the observed data $\mathbf{x}_0$ were generated using some fixed unknown ξ_0 and $\mathbf{u}_0$, i.e., $\mathbf{x}_0 = G(\xi_0, \mathbf{u}_0)$. The information that the solution of the equation $\mathbf{x}_0 = G(\xi, \mathbf{U})$ 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 $\mathbf{x}_0 = G(\cdot\,, \mathbf{u})$ 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. However, 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 the Borel paradox Casella and Berger (2002, Section 4.9.3).

Careful evaluation of the event "there is at least one ξ solving the equation $\mathbf{x}_0 = G(\xi, \mathbf{u})$" reveals that it has probability zero only if the probability of generating the realized data is zero. Hence, the Borel paradox is not an issue when defining generalized fiducial distribution for discrete models. Taking this observation a step further, notice that any data that a statistician can come into contact with has been rounded off in some manner, e.g., by a measuring instrument or by storage on a computer. Mathematically speaking, we do not know the exact realized value $\mathbf{X} = \mathbf{x}_0$. Instead we only observe an occurrence of an event $\{\mathbf{X} \in A_{\mathbf{x}_0}\}$, for some multivariate interval $A_{\mathbf{x}_0} = [\mathbf{a}, \mathbf{b})$ containing $\mathbf{x}_0$ and satisfying $P_{\xi_0}(\mathbf{X}^\star \in A_{\mathbf{x}_0}) > 0$, where $\mathbf{X}^\star$ is an independent copy of $\mathbf{X}$. To demonstrate this, if the exact unobserved value of the random vector $\mathbf{X}$ were $\mathbf{x}_0 = (\pi, e, 1.28)$ and, due to instrument precision, all the values were rounded down to one decimal place, our observation would be the knowledge that the event $\{\mathbf{X} \in [3.1, 3.2) \times [2.7, 2.8) \times [1.2, 1.3)\}$ happened.

Replacing the event of probability zero $\{\mathbf{X} = \mathbf{x}_0\}$ with the event of positive probability $\{\mathbf{X} \in A_{\mathbf{x}_0}\}$ removes non-uniqueness due to the Borel paradox in the definition of generalized fiducial distribution. This is done without any loss of information as only the occurrence of the event $\{\mathbf{X} \in A_{\mathbf{x}_0}\}$ is known to us and the interval $A_{\mathbf{x}_0}$ is a member of some fixed partition of $\mathbb{R}^n$ determined by the measuring instrument or computer precision. Using the language of σ-algebras, discretization is accommodated by restriction of the the Borel σ-algebra to a sub-σ-algebra generated by the countable partition $\{A_i\}$ whose events have positive P_{ξ_0} probability.

In situations where exact, non-discretized data is available, we propose to define the generalized fiducial distribution as a limit offering an attractive resolution of the Borel paradox. To this end, we first study the limit of the generalized fiducial distribution for a fixed sample size of jointly absolutely continuous random variables under general conditions as the precision of the discretization increases, $(\mathbf{b} - \mathbf{a}) \to 0$. We derive the limit in a closed form and show that it does not suffer from non-uniqueness due to multiple solutions of $\mathbf{x}_0 = G(\xi, \mathbf{u})$. Indeed the limiting distribution is the conditional distribution conditioned on the limit of the σ-algebras generated by the discretizations.

Second, we study the limit of the generalized fiducial distribution for i.i.d. data as the sample size goes to infinity and the discretization of the data remains fixed. We show that under very mild conditions the generalized fiducial distribution always leads to asymptotically correct inference. Here we evaluate the quality of an inference procedure in the repeated sample frequentist sense. To do this we effectively prove a Bernstein-von Mises theorem for generalized fiducial distributions and show that the effect of the particular selection of one of the ξ solving $G(\xi, \mathbf{u}) \in A_{\mathbf{x}_0}$ is of a second order as the sample size increases. Our result greatly relaxes the conditions under which the asymptotic correctness of generalized fiducial distribution has been previously proved.

The third source of non-uniqueness in the definition of the generalized fiducial distribution is due to the choice of structural equation (1.2). In particular, two different structural equations resulting in the same sampling distribution for data can lead to a different generalized fiducial distribution. While we do not fully resolve this issue, we offer some practical suggestions and comments on this topic.

The rest of the paper is organized as follows. In Section 2 we provide a rigorous definition of the generalized fiducial distribution. Section 3 studies the limit of the fiducial distribution as the precision of the data increases. Section 4 explores large sample asymptotics for the generalized fiducial inference under the presence of discretized data. Thoughts on the choice of structural equation are offered in Section 5. Section 5 concludes.

2. Generalized Fiducial Inference

We closely follow the definition of generalized fiducial distribution found in Hannig (2009) with a small modification to allow for discretized data. In order to avoid repeating the same arguments, we refer the reader to Section 4 of Hannig (2009) for a more detailed development.

Let $\mathbf{X} \in \mathbb{R}^n$ be a random vector with a distribution indexed by a parameter $\xi \in \Xi$. Assume that the data generating mechanism for $\mathbf{X}$ is expressed by (1.2) where G is a jointly measurable function and $\mathbf{U}$ is a random variable or vector with a completely known distribution independent of any parameters. We define for each measurable set $A \subset \mathbb{R}^n$ and all $\mathbf{u}$ a set-valued function

$$Q(A, \mathbf{u}) = \{\xi : \mathbf{G}(\xi, \mathbf{u}) \in A\}. \tag{2.1}$$

The function $Q(A, \mathbf{u})$ is the inverse image of the function $\mathbf{G}$ for fixed $\mathbf{u}$.

Next, we select a possibly random point out of each inverse image $Q(A, \mathbf{u})$. Following Section 4 of Hannig (2009), let $\{V(S)\}_{S \in \mathcal{B}^p}$ be a collection of random elements each with support $\bar{S}$. Since we will use $V(Q(A, \mathbf{u}))$ in the definition of the generalized fiducial distribution, the random elements $\{V(S)\}_{S \in \mathcal{B}^p}$ should be selected to be as uninformative as possible. A good default choice is a selection that maximizes the dispersion of the parameters of interest. For example if $S = (a, b) \subset \mathbb{R}$, $V((a, b))$ selects one of the endpoint a, b at random maximizing the variance of $V((a, b))$, or if S is a polyhedron and the parameters of interest are a subset of all parameters, $V(S)$ first projects the polyhedron on the subspace of the parameters of interest and then selects one of the vertices of the projection at random maximizing the determinant of the relevant covariance matrix. Simulation studies in Section 6 of Hannig (2009) and E, Hannig, and Iyer

(2009) examine the effects of the choice of $\{V(S)\}_{S\in\mathcal{B}^p}$ on frequentist behavior of generalized fiducial distribution arriving to a similar recommendation.

Assume that our data were generated by (1.2) using some true unknown parameter value ξ_0 and, instead of observing the exact realized value $\mathbf{X} = \mathbf{x}_0$, we only observe the event that the sample values lie in some measurable set $\{\mathbf{X} \in A_{\mathbf{x}_0}\}$, where $\mathbf{x}_0 \in A_{\mathbf{x}_0}$ is a member of a partition of $\mathbb{R}^n$. In addition to the information given to us by observing $\{\mathbf{X} \in A_{\mathbf{x}_0}\}$, we also know that the true values of ξ_0 and $\mathbf{u}_0$ satisfy $G(\xi_0, \mathbf{u}_0) \in A_{\mathbf{x}_0}$. Using the argument immediately preceding Equation (4.3) in Hannig (2009), we define a generalized fiducial distribution for ξ as

$$V(Q(A_{\mathbf{x}_0}, \mathbf{U}^\star)) \mid \{Q(A_{\mathbf{x}_0}, \mathbf{U}^\star) \neq \emptyset\}, \tag{2.2}$$

where $\mathbf{U}^\star$ is an independent copy of $\mathbf{U}$. The conditional distribution in (2.2) is well-defined provided that $P(Q(A_{\mathbf{x}_0}, \mathbf{U}^\star) \neq \emptyset) > 0$, which is the case as soon as $P(\mathbf{X}^\star = \mathbf{G}(\xi_0, \mathbf{U}^\star) \in A_{\mathbf{x}_0}) > 0$, since $\{\mathbf{u} : \mathbf{G}(\xi_0, \mathbf{u}) \in A_{\mathbf{x}_0}\} \subset \{\mathbf{u} : Q(A_{\mathbf{x}_0}, \mathbf{u}) \neq \emptyset\}$. Otherwise, additional care is needed to interpret the conditional distribution. We provide such an interpretation in Section 3. A generalized fiducial distribution for a subset θ of the parameter vector ξ is obtained through marginalization of the distributions in (2.2) Hannig (2009, Equation (4.4)). Finally notice that for exactly observed data we have $A_{\mathbf{x}_0} = \{\mathbf{x}_0\}$ and (2.2) is the same as (4.3) in Hannig (2009).

We prove that the effect of the particular choice of $\{V(S)\}_{S\in\mathcal{B}^p}$ disappears asymptotically. In order to simplify some of the notation in the proofs we modify the generalized fiducial distribution by having it defined as a probability distribution on the set of all subsets 2^Θ;

$$Q(A_{\mathbf{x}_0}, \mathbf{U}^\star) \mid \{Q(A_{\mathbf{x}_0}, \mathbf{U}^\star) \neq \emptyset\}. \tag{2.3}$$

The object defined in (2.3) is a random set of parameters (such as an interval or a polygon) with distribution conditioned on the set being non-empty. If there is no danger of misunderstanding, we call the modified generalized fiducial distribution of (2.3) also a generalized fiducial distribution.

Examples 1 and 2 of Section 4 of Hannig (2009) provide simple illustrations of the definition of generalized fiducial distribution for exactly observed data. An example provides a slightly more complicated illustration of the definition of a generalized fiducial distribution for discretized continuous data.

Example 1. Suppose $\mathbf{U} = (U_1, \ldots, U_n)$, where U_i are i.i.d. $N(0, 1)$ and

$$\mathbf{X} = (X_1, \ldots, X_n) = G(\mu, \mathbf{U}) = (\mu + U_1, \ldots, \mu + U_n)$$

for some $\mu \in \mathbb{R}$, so the X_is are i.i.d. $N(\mu, 1)$. We observe a discretized realization of $\mathbf{X}$, i.e., the event $\{\mathbf{X} \in (\mathbf{a}, \mathbf{b})\}$, where $\mathbf{a} = (a_1, \ldots, a_n)$, $\mathbf{b} = (b_1, \ldots, b_n)$ and $(\mathbf{a}, \mathbf{b})$ is an n-dimensional cube determined by the way the data is rounded off at the measuring device.

If $n = 1$ then $Q((a, b), u) = (a-u, b-u)$ and $P_{\xi_0}(Q((a, b), U^\star) \neq \emptyset) = 1$. Thus following (2.3), the modified generalized fiducial distribution is the distribution of the random interval $(a - U^\star, b - U^\star)$ where $U^\star \sim N(0, 1)$, independent of the data.

If $n > 1$, take $L(\mathbf{a}, \mathbf{u}) = \max_i\{a_i - u_i\}$ and $R(\mathbf{b}, \mathbf{u}) = \min_j\{b_j - u_j\}$. The inverse image is

$$Q((\mathbf{a}, \mathbf{b}), \mathbf{u}) = \begin{cases} (L(\mathbf{a}, \mathbf{u}), R(\mathbf{b}, \mathbf{u})) & \text{if } L(\mathbf{a}, \mathbf{u}) < R(\mathbf{b}, \mathbf{u}), \\ \emptyset & \text{otherwise.} \end{cases}$$

Using $\Phi(x)$ and $\varphi(x)$ for the distribution function and density of $N(0, 1)$, respectively, we compute for constants l, r,

$$\begin{aligned} P(L(\mathbf{a}, \mathbf{U}^\star) \leq l, r < R(\mathbf{b}, \mathbf{U}^\star)) &= P(a_i - l \leq U_i^\star < b_i - r, \text{for all } i) \\ &= \prod_{i=1}^{n} (\Phi(b_i - r) - \Phi(a_i - l))_+ \end{aligned} \tag{2.4}$$

Notice that the probability in (2.4) is not zero if and only if $b_i - r > a_i - l$, for all $i = 1, \ldots, n$, which is equivalent to $\Delta > r - l$ with $\Delta = \min_i\{b_i - a_i\}$. The joint density $L(\mathbf{a}, \mathbf{U}^\star), R(\mathbf{b}, \mathbf{U}^\star)$ is computed by taking derivatives, and the modified generalized fiducial distribution for μ, (2.3), is the distribution of the random interval (L, R), where the joint density $f_{LR}(l, r)$ is

$$\frac{\sum_{i \neq j} \Big(\varphi(a_i - l)\varphi(b_j - r) \prod_{k \notin \{i,j\}} (\Phi(b_k - r) - \Phi(a_k - l))\Big) I_{\{l < r < l + \Delta\}}}{\int_0^\infty \int_{l'}^{\Delta + l'} \sum_{i \neq j} \Big(\varphi(a_i - l')\varphi(b_j - r') \prod_{k \notin \{i,j\}} (\Phi(b_k - r') - \Phi(a_k - l'))\Big) \, dr' dl'}.$$

The generalized fiducial distribution for μ, (2.2), is obtained by selecting a point inside of the interval $[L, R]$. A reasonable default is to take $V((L, R)) = L$ with probability .5 and $V((L, R)) = R$ with probability .5, independently of everything else.

3. Increasing Precision Asymptotics

In this section we discuss the behavior of the generalized fiducial distribution as we increase the precision of the measurements. This provides a definition of the generalized fiducial distribution for exactly observed observations. Such asymptotic considerations are not relevant for discrete distributions, and therefore we turn our attention to distributions that are absolutely continuous with respect to Lebesgue measure.

Let us now state the assumptions of Theorem 1; these are weaker than the assumptions stated in Section 4.1 of Hannig (2009). In particular the current assumptions apply to a wider selection of models than just the i.i.d. sequences covered in Hannig (2009). For examples, see the end of this section.

Assume that the realized value of $\mathbf{X}$, generated using some true unknown parameter value ξ_0, is $\mathbf{x}_0$. Suppose that the parameter of interest $\xi_0 \in \Xi \subset \mathbb{R}^p$ is p-dimensional. Recall (1.2), and assume that $\mathbf{U} \in \mathbb{R}^n$ is an absolutely continuous random vector with a joint density $f_{\mathbf{U}}(\mathbf{u})$, defined with respect to Lebesgue measure on $\mathbb{R}^n$, continuous on its support. Write $\mathbf{G} = (g_1, \ldots, g_n)$ so that $X_i = g_i(\xi, \mathbf{U})$ for $i = 1, \ldots, n$. Assume that for each fixed $\xi \in \Xi$ the function $\mathbf{G}(\xi, \boldsymbol{.}\,)$ is one-to-one and continuously differentiable, denoting the inverse by $\mathbf{G}^{-1}(\mathbf{x}, \xi)$. Using the Jacobian transformation, the density of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x} \mid \xi) = f_{\mathbf{U}}\left(\mathbf{G}^{-1}(\mathbf{x}, \xi)\right) \left| \det\left(\frac{\mathbf{d}}{\mathbf{dx}} \mathbf{G}^{-1}(\mathbf{x}, \xi) \right) \right|. \tag{3.1}$$

For all p-tuples of indexes $\mathbf{i} = (1 \le i_1 < \cdots < i_p \le n) \subset \{1, \ldots, n\}$ we denote the list of unused indexes by $\mathbf{i}^{\complement} = \{1, \ldots, n\} \setminus \mathbf{i}$, the collection of variables indexed by $\mathbf{i}$ by $\mathbf{x_i} = (x_{i_1}, \ldots, x_{i_p})$, and its complement by $\mathbf{x}_{\mathbf{i}^{\complement}} = (x_i : i \in \mathbf{i}^{\complement})$. Assume that there is an open neighborhood $\mathcal{B}(\mathbf{x}_0)$ and a measurable sets $\mathcal{U}_{\mathbf{i}}$, $P(\mathbf{U_i} \in \mathcal{U}) > 0$, such that, for all $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{B}(\mathbf{x}_0)$ and for all p-tuples of indexes $\mathbf{i}$, the function $\mathbf{G}^{-1}\left((\mathbf{x_i}, \boldsymbol{.}\,), \boldsymbol{.}\,\right)$, viewed as a function of ξ and $\mathbf{x}_{\mathbf{i}^{\complement}}$, is one-to-one and differentiable onto $\mathcal{U}_{\mathbf{i}}$. Thus, the density of $(\xi, \mathbf{X}_{\mathbf{i}^{\complement}})$ is

$$f_{\xi \mathbf{X}_{\mathbf{i}^{\complement}}}(\xi, \mathbf{x}_{\mathbf{i}^{\complement}} | \mathbf{x_i}) = f_{\mathbf{U}}\left(\mathbf{G}^{-1}(\mathbf{x}, \xi)\right) \left| \det\left(\frac{\mathbf{d}}{\mathbf{d}(\xi, \mathbf{x}_{\mathbf{i}^{\complement}})} \mathbf{G}^{-1}(\mathbf{x}, \xi) \right) \right|. \tag{3.2}$$

Here $\frac{d}{\mathbf{d}(\xi, \mathbf{x}_{\mathbf{i}^{\complement}})} \mathbf{G}^{-1}(\mathbf{x}, \xi)$ stands for the Jacobian matrix computed with respect to all parameters ξ and all observations $\mathbf{x}_{\mathbf{i}^{\complement}}$. It follows that for any fixed ξ the function $f_{\xi \mathbf{X}_{\mathbf{i}^{\complement}}}(\xi, \mathbf{x}_{\mathbf{i}^{\complement}} | \mathbf{x_i})$ is continuous in $\mathbf{x} = (\mathbf{x_i}, \mathbf{x}_{\mathbf{i}^{\complement}})$. Assume additionally that the marginal density $\int_{\Xi} f_{\xi \mathbf{X}_{\mathbf{i}^{\complement}}}(\xi, \mathbf{x}_{\mathbf{i}^{\complement}} | \mathbf{x_i})\, d\xi$ is continuous in $\mathbf{x} = (\mathbf{x_i}, \mathbf{x}_{\mathbf{i}^{\complement}})$.

Finally, consider a sequence of discretizations of $\mathbf{x}_0 = (x_{1,0}, \ldots, x_{n,0})$. In particular, for each $m = 1, 2, \ldots$, each coordinate $x_{0,i} \in (a_{i,m}, b_{i,m})$ for all $i = 1, \ldots, n$. Let $\mathbf{a}_m = (a_{1,m}, \ldots, a_{n,m})$, $\mathbf{b}_m = (b_{1,m}, \ldots, b_{n,m})$, and assume that for all $m = 1, 2, \ldots$ the probability $P_{\xi_0}(\mathbf{X} \in (\mathbf{a}_m, \mathbf{b}_m)) > 0$, so that the conditional distributions in (2.3) are uniquely defined.

Theorem 1. *Under the assumptions of this section, consider a sequence of p-dimensional intervals $(\mathbf{a}_1, \mathbf{b}_1) \supset (\mathbf{a}_2, \mathbf{b}_2) \supset \cdots$ and numbers $c_m \uparrow \infty$ such that*

$\bigcap_m(\mathbf{a}_m, \mathbf{b}_m) = \{\mathbf{x}_0\}$ *and* $c_m(b_{m,i} - a_{m,i}) \to w_i > 0$ *for all* $i = 1, \ldots, n$. *Then the modified generalized fiducial distribution*

$$Q((\mathbf{a}_m, \mathbf{b}_m), U^\star) \mid \{Q((\mathbf{a}_m, \mathbf{b}_m), U^\star) \neq \emptyset\} \tag{3.3}$$

converges weakly to a singleton that has an absolutely continuous distribution with density

$$r(\xi) = \frac{f_{\mathbf{X}}(\mathbf{x}_0|\xi)J(\mathbf{x}_0, \xi)}{\int_\Xi f_{\mathbf{X}}(\mathbf{x}_0|\xi')J(\mathbf{x}_0, \xi')\, d\xi'}, \tag{3.4}$$

where $f_{\mathbf{X}}(\mathbf{x}_0|\xi)$ *is the likelihood function and*

$$J(\mathbf{x}, \xi) = \sum_{\substack{\mathbf{i}=(i_1,\ldots,i_p) \\ 1 \le i_1 < \cdots < i_p \le n}} \frac{1}{w_{i_1} \ldots w_{i_p}} \left| \frac{\det\left(\frac{\mathbf{d}}{\mathbf{d}(\xi, \mathbf{x}_{\mathbf{i}^\complement})} \mathbf{G}^{-1}(\mathbf{x}, \xi)\right)}{\det\left(\frac{\mathbf{d}}{\mathbf{dx}} \mathbf{G}^{-1}(\mathbf{x}, \xi)\right)} \right|. \tag{3.5}$$

As discussed in Hannig (2009), Section 4, Remark 1, there are three sources of non-uniqueness in the definition of the fiducial distribution: the choice of structural equation, the Borel paradox if $P_{\xi_0}(Q(A, U^\star) \neq \emptyset) = 0$, and the choice of a particular value in $Q(A, U^\star)$ if it contains more than one element. Theorem 1 gives a reasonable, consistent way of resolving the non-uniqueness due to the last two issues for large class of models.

In particular, the limit of conditional distributions (3.3) contains only one element with probability 1, and the non-uniqueness due to the choice of a particular value in $Q(A, U^\star)$ is therefore not present in the limit. Moreover, since $\lim_{m\to\infty}(\mathbf{a}_m, \mathbf{b}_m) = \{\mathbf{x}_0\}$, the limiting probability density (3.4) can be taken as an appealing implementation of the conditional distribution (2.2) with $A = \{\mathbf{x}_0\}$, resolving non-uniqueness due to the Borel paradox. Finally, the w_i are fully determined by the relative limiting size of the discretization. For example we have $w_1 = \cdots = w_n = 1$ if the observed data is discretized to the same precision and recorded on the same scale, such as in the case of i.i.d. observations measured by the same instrument.

A proposition shows that the limiting generalized fiducial distribution in (3.4) and (3.5) is invariant under smooth reparametrizations. This is a desirable property similar to the invariance of the posterior computed using the Jeffreys prior.

Proposition 1. *Let* $\xi = \phi(\eta)$ *be a one-to-one, continuously differentiable function onto the parameter space* Ξ. *Let* $r(\xi)$ *be the generalized fiducial distribution computed from* $\mathbf{X} = \mathbf{G}(\xi, \mathbf{U})$ *using* (3.4) *and* (3.5), *and* $\tilde{r}(\eta)$ *the generalized fiducial distribution computed from* $\mathbf{X} = \mathbf{G}(\phi(\eta), \mathbf{U})$ *using* (3.4) *and* (3.5). *Then for each measurable set* $A \subset \Xi$

$$\int_A r(\xi)\, d\xi = \int_{\phi^{-1}(A)} \tilde{r}(\eta)\, d\eta.$$

Proof. The multivariate chain rule reveals, after simplification of the second determinant that

$$\det\left(\frac{\mathbf{d}}{\mathbf{d}(\eta,\mathbf{x}_{\mathbf{i}^\complement})}\mathbf{G}^{-1}(\mathbf{x},\phi(\eta))\right)=\det\left(\frac{\mathbf{d}}{\mathbf{d}(\phi(\eta),\mathbf{x}_{\mathbf{i}^\complement})}\mathbf{G}^{-1}(\mathbf{x},\phi(\eta))\right)\det\left(\frac{\mathbf{d}}{\mathbf{d}\eta}\phi(\eta)\right).$$

By the usual Jacobian transformation

$$\begin{aligned}
&\int_A f_{\mathbf{X}}(\mathbf{x}_0|\xi)\left|\frac{\det\left(\frac{\mathbf{d}}{\mathbf{d}(\xi,\mathbf{x}_{\mathbf{i}^\complement})}\mathbf{G}^{-1}(\mathbf{x},\xi)\right)}{\det\left(\frac{\mathbf{d}}{\mathbf{dx}}\mathbf{G}^{-1}(\mathbf{x},\xi)\right)}\right|d\xi\\
&\quad=\int_{\phi^{-1}(A)} f_{\mathbf{X}}(\mathbf{x}_0|\phi(\eta))\left|\frac{\det\left(\frac{\mathbf{d}}{\mathbf{d}(\phi(\eta),\mathbf{x}_{\mathbf{i}^\complement})}\mathbf{G}^{-1}(\mathbf{x},\phi(\eta))\right)}{\det\left(\frac{\mathbf{d}}{\mathbf{dx}}\mathbf{G}^{-1}(\mathbf{x},\phi(\eta))\right)}\right|\left|\det\left(\frac{\mathbf{d}}{\mathbf{d}\eta}\phi(\eta)\right)\right|d\eta\\
&\quad=\int_{\phi^{-1}(A)} f_{\mathbf{X}}(\mathbf{x}_0|\phi(\eta))\left|\frac{\det\left(\frac{\mathbf{d}}{\mathbf{d}(\eta,\mathbf{x}_{\mathbf{i}^\complement})}\mathbf{G}^{-1}(\mathbf{x},\phi(\eta))\right)}{\det\left(\frac{\mathbf{d}}{\mathbf{dx}}\mathbf{G}^{-1}(\mathbf{x},\phi(\eta))\right)}\right|d\eta.
\end{aligned}$$

The statement now follows by simple algebra.

If the observations are from an i.i.d. univariate absolutely continuous distribution, we can choose a particular structural equation (1.2) that recovers Fisher's original definition of fiducial distribution. To this end, with $F(x,\xi)$ and $f(x,\xi)$ the distribution and density functions, respectively, define the usual pseudo-inverse $F^{-1}(\xi,u)=\inf_x\{F(x,\xi)\geq u)$ and use the structural equation

$$X_i=F^{-1}(\xi,U_i),\quad i=1,\ldots,n, \tag{3.6}$$

where U_i are i.i.d. $U(0,1)$. If, additionally, the assumptions of Theorem 1 are satisfied, the inverse of the structural equation $\mathbf{u}=\mathbf{G}^{-1}(\mathbf{x},\xi)$ is

$$u_i=F(x_i,\xi),\quad i=1,\ldots,n,$$

and the generalized fiducial distribution is (3.4), with (3.5) simplified to

$$J(\mathbf{x},\xi)=\sum_{\substack{\mathbf{i}=(i_1,\ldots,i_p)\\1\leq i_1<\cdots<i_p\leq n}}\left|\frac{\det\left(\frac{\mathbf{d}}{\mathbf{d}\xi}\left(F(x_{i_1},\xi),\ldots,F(x_{i_p},\xi)\right)\right)}{f(x_{i_1},\xi)\cdots f(x_{i,p},\xi)}\right|. \tag{3.7}$$

If $n=p=1$, (3.4) and (3.7) become (1.1). Similarly if $n\geq p=1$, (3.4) and (3.7) agree with the proposal of Dempster (1963).

We remark that (3.4) and (3.5) agree with, validate, and generalize a heuristically motivated proposal of Hannig (2009), Section 4.1, which uses a particularly simple idea to implement the conditional distribution in the definition of generalized fiducial distribution. This is an unexpected result because the heuristically motivated proposal is based on picking p equations at random, solving for the parameters using the selected equations, and conditioning on the rest of the equations, while the result presented here is a consequence of the geometry of the random sets used in the definition of the generalized fiducial distribution.

Hannig (2009) relates, in Section 4.2, (3.4) and (3.5) to Lindley (1958), see also Dempster (1963). It is also of interest that, in the same section, Hannig (2009) notices that the function $J(\mathbf{x}, \xi)$ can be viewed as a U-statistic estimator of $\pi(\xi) = E_{\xi_0} J(\mathbf{X}^*, \xi)$, where $\mathbf{X}^*$ is an independent copy of the data, giving the generalized fiducial distribution an empirical Bayes interpretation.

We remark that the $J(\mathbf{x}, \xi)$ of (3.7) is related to the data dependent priors proposed by Fraser, Reid, Marras, and Yi (2010). Consider a matrix

$$V(\mathbf{x}, \xi) = \begin{pmatrix} \frac{\frac{\mathbf{d}}{\mathbf{d}\xi} F(x_1, \xi)}{f(x_1, \xi)} \\ \vdots \\ \frac{\frac{\mathbf{d}}{\mathbf{d}\xi} F(x_n, \xi)}{f(x_n, \xi)} \end{pmatrix}.$$

The solution in (3.7), derived as the limit of the generalized fiducial distribution for discretized data, obtains the data dependent default prior $J(\mathbf{x}, \xi)$ as a sum of all possible absolute values of determinants of $p \times p$ matrices obtained by selecting p rows from $V(\mathbf{x}, \xi)$. Alternatively, Fraser, Reid, Marras, and Yi (2010) consider their data dependent prior as $|\det(A(\mathbf{x}, \xi)V(\mathbf{x}, \xi))|^q$, where $q > 0$ and $A(\mathbf{x}, \xi)$ is a suitable matrix. Fraser, Reid, Marras, and Yi (2010) propose several choices of $A(\mathbf{x}, \xi)$ but, as a reasonable default motivated by maximum likelihood ideas, recommend $q = 1$ and $A(\mathbf{x}, \xi) = (\frac{\mathbf{d}^2}{\mathbf{d}\xi \mathbf{d}\xi'} l(\mathbf{x}, \xi))^{-1/2} \frac{\mathbf{d}^2}{\mathbf{d}\xi \mathbf{d}\mathbf{x}} l(\mathbf{x}, \xi)$, where $l(\mathbf{x}, \xi)$ is the log likelihood of the data. The drawback of this proposal is that it requires the existence of a second derivative of the log-likelihood. If the log-likelihood is not differentiable, they recommend $q = 1/2$ and $A(\mathbf{x}, \xi) = V(\mathbf{x}, \xi)^\top$. They do not provide any simulation studies that would exhibit small sample performance.

To conclude this section we consider two examples.

Example 2. Let $X_1, \ldots, X_n$ be i.i.d. $U(\theta, \theta^2)$ random variables, $\theta > 1$. Using the inverse distribution function for a structural equation we get

$$X_i = \theta(\theta - 1)U_i + \theta, \quad i = 1, \ldots, n$$

with U_i i.i.d. $U(0, 1)$. Using the limit in (3.3) and (3.7) we get the generalized fiducial density

$$r(\theta) \propto \frac{I_{(x_{(n)}^{1/2}, x_{(1)})}(\theta)}{(\theta(\theta-1))^n} \cdot \frac{\sum_{i=1}^n x_i(2\theta - 1) - n\theta^2}{\theta(\theta - 1)}, \tag{3.8}$$

where the first term on the right side of (3.8) is the likelihood and the second term is the Jacobian factor in (3.7).

We performed a limited simulation study to validate the frequentist performance of the confidence intervals based on the generalized fiducial distribution (3.8). We used $\theta = 1.01, 1.5, 2, 10, 50, 250$ and sample sizes $n = 1$, 2, 3, 5, 10, 20, 100. The simulation results show that confidence intervals based on the generalized fiducial distribution have nearly exact frequentist coverage for all parameter combinations and all confidence levels. Moreover, the expected length of the proposed 95% equal tailed confidence intervals based on (3.8) was slightly shorter than the 95% intervals based on the reference prior solution of Berger, Bernardo, and Sun (2009) and the proposal of Fraser, Reid, Marras, and Yi (2010). The details of the simulation study are available from the author upon request.

Example 3. Consider the Gaussian AR(1) model. The usual model formulation $X_i = aX_{i-1} + Z_i$, with Z_i i.i.d. $N(0, \sigma^2)$, can be reexpressed as the structural equation

$$X_i = a^i x_0 + \sigma \sum_{j=1}^{i} a^{i-j} U_j, \quad i = 1, \ldots, n,$$

with parameters $\xi = (a, \sigma, x_0)$ and random component $\mathbf{U} = (U_1, \ldots, U_n)$, where the U_i are i.i.d. $N(0, 1)$. The inverse of the structural equation $\mathbf{u} = \mathbf{G}^{-1}(\mathbf{x}, \xi)$ is

$$u_i = \frac{x_i - ax_{i-1}}{\sigma}, \quad i = 1, \ldots, n.$$

The generalized fiducial distribution in (3.4) is

$$r(\xi) \propto (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{\sum_{i=1}^{n}(x_i - ax_{i-1})^2}{2\sigma^2}\right) J(\mathbf{x}, \xi),$$

with $J(\mathbf{x}, \xi)$ given by (3.5). To compute $J(\mathbf{x}, \xi)$, notice that the Jacobian matrix in the denominator is triangular and therefore the Jacobian is

$$\det\left(\frac{\mathbf{d}}{\mathbf{dx}}\mathbf{G}^{-1}(\mathbf{x}, \xi)\right) = \sigma^{-n}.$$

The Jacobians in the numerator are more complicated but careful algebra reveals that if $\mathbf{i} = (i, j, k)$, $1 \le i < j < k \le n$,

$$\det\left(\frac{d}{\mathbf{d}(\xi, \mathbf{x}_{\mathbf{i}^{\complement}})}\mathbf{G}^{-1}(\mathbf{x}, \xi)\right) = \sum_{l=1}^{k-i-1} \frac{(-1)^{i+l-1} a^{l+i-1} \left(x_{\max\{j-l,i\}} x_{\min\{k,k+j-l-i\}} - x_j x_{k-l}\right)}{\sigma^{n+1}}.$$

Recall that $\frac{d}{\mathbf{d}(\xi,\mathbf{x}_{\mathbf{i}\mathsf{C}})}\mathbf{G}^{-1}(\mathbf{x},\xi)$ stands for the Jacobian matrix computed by taking derivatives with respect to the parameters (a,σ,x_0) and x_l, $l \neq i,j,k$. Again, $w_i = 1$ and $J(\mathbf{x},\xi)$ is obtained by simple algebra.

Numerical studies reveal an interesting property of the generalized fiducial distribution for the Gaussian AR(1) model. If the observed X_i are stationary, the marginal generalized fiducial distribution for (a,σ) is bimodal with one mode near the true values (a_0,σ_0) and the other near $(a_0^{-1},\sigma_0|a_0|^{-1})$. The existence and location of the second mode is intriguing, given that the second mode is near the parameters of the same time series run backwards in time $\tilde{X}_i = X_{n-i}$. To explain this, recall that the distribution of a stationary Gaussian AR(1) series is the same as the distribution of the time reversed stationary time series, as both have the same covariance function. The existence of the two modes in the generalized fiducial density therefore correctly reflects the fact that we cannot distinguish causal and non-causal stationary time series based on observations only. Since the time series is stationary, we might feel at the first glance that this non-causal bump is superfluous. However, the direction of the time series is important for predicting the starting value $\hat{X}$ and the joint generalized fiducial distribution correctly recognizes the non-identifiability in the time direction. This is all the more exciting because the likelihood function itself has only one mode near the causal values, $|a_0| < 1$.

We remark that based on our simulations, if the true $|a_0| < 1$ and we assume that the starting value $\hat{X}$ is observed, the corresponding generalized fiducial distribution does not have the second non-causal mode. Similarly, if the observed time series is far from stationary, both the likelihood and the marginal generalized fiducial distribution for (a,σ) is unimodal with mode near the true value (a_0,σ_0), regardless of whether the time series is causal or not.

4. Increasing Sample Size Asymptotics

In this section we look at the behavior of the generalized fiducial distribution for i.i.d. random variables as the number of observations increases and observational discretization remains fixed. The conditions stated here are weak and easy to verify. They are formulated in terms of the distribution function, and only the existence and continuity of the first partial derivative with respect to the parameters is assumed. Also, unlike in Section 3 where only the intervals including the fixed realized data $\mathbf{x}_0$ are considered, here we are investigating repeated sampling performance and need to know all the members of the partition discretizing the real line.

Assume the structural equation (3.6),

$$X_i = F^{-1}(\xi, U_i), \quad i = 1,\ldots,n,$$

where the X_i are random variables, $\xi \in \Xi \subset \mathbb{R}^p$ is a p-dimensional parameter, the U_i are i.i.d. $U(0,1)$ and $F^{-1}(\xi, u) = \inf_x\{F(x,\xi) \ge u\}$. We choose this structural equation, because it fits naturally into the structure of our proof and does not require introduction of additional assumptions. If another structural equation generating the same sampling distribution of the data were chosen, additional assumptions on this structural equation would be required.

Assume that $\mathbb{R}$ is partitioned into the fixed intervals $(-\infty, a_1], (a_1, a_2], \ldots,$ (a_k, ∞) with $a_0 = -\infty,\ a_{k+1} = \infty$. The values of X_i are observed only up to the resolution of the grid, i.e., we observe $\mathbf{k} = (k_1, \ldots, k_n)$ so that $x_i \in (a_{k_i}, a_{k_i+1}]$, or equivalently $\mathbf{x} \in (\mathbf{a_k}, \mathbf{a_{k+1}}]$ with $\mathbf{a_k} = (a_{k_1}, \ldots, a_{k_n})$.

Assume that $k \ge p$, $p_j(\xi) = P\left(X \in (a_j, a_{j+1}]\right) > 0$ for all $j = 0, \ldots, k$ and all ξ. Assume $F(x,\xi)$ is continuously differentiable in ξ for all $x \in \{a_1, \ldots, a_k\}$. Additionally, assume for all $\mathbf{j} = (j_1 < \cdots < j_p) \subset \{1, \ldots, k\}$ that $(F(a_{j_1}, \xi), \ldots,$ $F(a_{j_p}, \xi)) = (u_1, \ldots, u_p)$ as a function in ξ with $u_1 < \cdots < u_p$, is one-to-one and the Jacobian

$$\det\left(\frac{\mathbf{d}(F(a_{j_1}, \xi), \ldots, F(a_{j_p}, \xi))}{\mathbf{d}\xi}\right) \ne 0.$$

Finally, let $\mathcal{R}_\xi$ be a random variable having the generalized fiducial distribution (2.2).

Theorem 2. *Under the assumptions of this section, $\mathcal{R}_\xi$ has an asymptotically normal distribution and any confidence set based on $\mathcal{R}_\xi$ of a shape satisfying Assumption 3 of* Hannig (2009) *has asymptotically correct coverage regardless of the choice of $V(\,\boldsymbol{\cdot}\,)$.*

The proof of the theorem is relegated to Appendix B. We remark that, as a consequence of this result, the do-not-know probability in Dempster-Shafer calculus (Dempster (2008)) vanishes and does not influence inference for large n.

The need to properly account for uncertainty due to discretization of observations modeled by continuous random variables is of particular importance in the field of metrology. The problem of inference for the mean of discretized normal data has obtained some attention in the last decade. Frenkel and Kirkup (2005) and Cordero, Seckmeyer, and Labbe (2006) proposed a maximum likelihood based approach, Willink (2007) used an ad-hoc modification of the sample mean, and Hannig, Iyer, and Wang (2007) proposed a generalized fiducial solution. Witkowsky and Wimmer (2009) report a thorough simulation study comparing the coverage and expected length of approximate confidence intervals for various sample sizes and levels of discretization. They compare approximate confidence intervals based on the standard Student t, asymptotic distribution of the maximum likelihood, the proposal of Willink (2007), and the generalized fiducial distribution. Among them, the Student t interval ignores the discretization while the rest accounts for it. Witkowsky and Wimmer (2009) report that the Student t intervals work poorly if the discretization is coarse; this is not surprising as it ignores the discretization. The maximum likelihood based confidence intervals are reported not to maintain the stated coverage for small sample sizes making them unreliable. Both Willink (2007) and the generalized fiducial solution performed adequately in terms of maintaining the stated coverage with the generalized fiducial intervals having uniformly shorter average length than the interval of Willink (2007). Hannig, Iyer, and Wang (2007) report a smaller simulation study that also shows good small sample performance of the generalized fiducial inference based intervals. Theoreom 2 therefore complements the good small sample properties by providing the necessary theoretical backing for the use of the generalized fiducial inference in practice.

5. Comments on the Choice of Structural Equation

The definition (1.2) is kept very general in order to make it applicable to many statistical models. This also means that the same sampling distribution can be generated by different structural equations. For example, it is well-known that one can always find a function $\mathbf{G}$ so that the random vector $\mathbf{X} = \mathbf{G}(\xi, U)$ where U is a single $U(0,1)$. However, such a choice can lead to generalized fiducial distributions that are mathematically and computationally intractable.

If $\mathbf{X}$ is absolutely continuous, we recommend choosing a structural equation so that the limiting generalized fiducial distribution (3.4) and (3.5) in Theorem 1 can be used. In particular the dimension of $\mathbf{X}$ should be the same as the dimension of the random vector $\mathbf{U}$, the inverse of the function $\mathbf{u} = \mathbf{G}^{-1}(\mathbf{x}, \xi)$ should exist and be continuously differentiable in ξ and $\mathbf{x}$, and the distribution of $\mathbf{U}$ should be absolutely continuous with a known simple density, e.g., $\mathbf{U} = (U_1, \ldots, U_n)$ with U_i i.i.d. $U(0,1)$ or $N(0,1)$. This recommendation is based on the fact that we find these assumptions necessary to derive a tractable expression for the generalized fiducial distribution and to prove its asymptotic properties, c.f., Section 5 of Hannig (2009). Moreover, Proposition 1 establishes that the generalized fiducial distribution given by (3.4) and (3.5) is invariant under one-to-one continuously differentiable reparametrizations of the parameter vector ξ.

Identifiability considerations imply that the structural equation should be chosen so that, for all disjoint A_1, A_2 and all $\mathbf{u}$, the sets $Q(A_1, \mathbf{u})$ and $Q(A_2, \mathbf{u})$ at (2.1) are disjoint.

If $\mathbf{X} = (X_1, \ldots, X_n)$ are i.i.d. with a distribution function differentiable in the parameter vector ξ, and the number of data points n is much larger than the number of parameters p, then the result in Section 4 together with its proof strongly suggest using a structural equation based on the inverse distribution function (3.6) as a reasonable default. Additionally, if $\mathbf{X}$ is absolutely continuous the generalized fiducial distribution obtained (3.4) and (3.7) could be viewed as a direct generalization of Fisher's original definition (1.1).

It is known that different structural equations generating data with the same sampling distribution can lead to different generalized fiducial distributions. We assert that this non-uniqueness is to be welcomed rather than eschewed. Wilkinson (1977) argues that the fiducial distribution should depend on the choice of parameter of interest, and the same argument has also been made in connection with the choice of an objective prior in Bayesian inference (Efron (1986); Berger, Bernardo, and Sun (2009)). Similarly, we conjecture that any general theory on the choice of the structural equation cannot ignore the parameter of interest. We demonstrate this conjecture with an example.

Example 4. Consider the sequence of independent random variables $X_i \sim N(\mu_i, 1),\ i = 1, \ldots, n$. The parameter of interest is $\theta = \left(\sum_{i=1}^n \mu_i^2\right)^{1/2}$. The nuisance parameter is a point on the unit sphere $\eta = \mu/\theta$, where $\mu = (\mu_1, \ldots, \mu_n) = \theta\eta$. The naïve structural equation

$$X_i = \mu_i + Z_i, \quad Z_i \text{ i.i.d. N(0,1)}, \quad i = 1, \ldots, n, \tag{5.1}$$

leads to the fiducial distribution that is the same as the Bayesian posterior computed with respect to the flat prior that is known to have exact frequentist properties for every individual μ_i but very bad frequentist properties for the parameter of interest θ.

Guided by our interest in θ, we propose another structural equation that isolates θ in a part of the structural equation. Write $\mathbf{X} = (X_1, \ldots, X_n)$. We model $\|\mathbf{X}\| = \left(\sum_{i=1}^n X_i^2\right)^{1/2}$ and $\mathbf{X}/\|\mathbf{X}\|$ separately. First, $\|\mathbf{X}\| = F_n^{-1}(\theta, U_1)$, where and $U_1 \sim U(0,1)$ and F_n^{-1} is the square root of the inverse of the non-central chi-square distribution function with n degrees of freedom and non-centrality parameter $\theta^2/2$. Next, $\mathbf{X}/\|\mathbf{X}\| = \eta \circ H_n(\theta, U_1, U_2)$, where $H_n(\theta, U_1, U_2)$ is an appropriate function generating $\mathbf{X}/\|\mathbf{X}\|$ if $\eta = (1, 0 \ldots, 0)$ were the truth, $\circ$ is the rotation group operator on the unit sphere and $U_2 \sim U(0,1)$ independent of U_1. By combining these two expressions we get the structural equation

$$\mathbf{X} = F_n^{-1}(\theta, U_1) \cdot (\eta \circ H_n(\theta, U_1, U_2)) \,. \tag{5.2}$$

Notice, that for any observed $\mathbf{x} \neq 0$ and any fixed $u_1, u_2 \in (0,1)$, there is unique θ, η solving (5.2). Moreover, θ is the solution to $\|\mathbf{x}\| = F_n^{-1}(\theta, u_1)$ only, and the resulting generalized fiducial distribution is based entirely on the non-central chi-square portion of the structural equation. It is well-known that the generalized fiducial distribution for θ, derived from the structural equation

$\|\mathbf{x}\| = F_n^{-1}(\theta, U_1^\star)$, leads to confidence intervals with very good frequentist properties, see for example the Example 5 of Hannig, Iyer, and Patterson (2006). We conclude that if θ is the parameter of interest, the structural equation (5.2) is preferable to the naïve structural equation (5.1).

A similar issue arises in the objective Bayes literature. If the parameter of interest is θ the default prior is $\pi(\mu) = \|\mu\|^{-(p-1)} = \theta^{-(p-1)}$ and not the naïve default prior $\pi(\mu) = 1$ (Stein (1985); Tibshirani (1989)).

It would be desirable if one could start with the naïve structural equation and obtain the "correct" structural equation by some well defined process. Search for such a process is a topic of our future research. Some promising ideas in that direction can be found in Zhang and Liu (2011).

6. Conclusions

In this paper we have studied asymptotic properties of generalized fiducial distribution for discretely observed data. The use of discretized data is natural because all data is discretized due to instrument precision and computer storage. The limiting distribution of the generalized fiducial distribution of discretely observed data as the precision of the discretization increases is obtained and used to resolve an ambiguity in the definition of generalized fiducial distribution for exactly observed data. We also show that, under some mild conditions on the parametric model, the generalized fiducial distribution for discretized data leads to asymptotically correct inference.

This paper did not deal with the computational issues surrounding generalized fiducial inference. Typically, a numerical scheme, such as MCMC or Sequential MC, needs to be employed. For example, Hannig, Iyer, and Wang (2007) implement a modified Gibbs sampler and show that generalized fiducial inference for discretized normal data indeed has very good small sample statistical properties. More complicated computational schemes for generalized fiducial inference can be also found in Hannig and Lee (2009), Wandler and Hannig (2011), and elsewhere.

Finally, we remark that there are interesting connections between generalized fiducial distribution, the asymptotic theory of likelihood Davison, Fraser, and Reid (2006), the theory of second order ancillaries Fraser, Fraser, and Staicu (2010) and as discussed in Section 3, the data dependent prior of Fraser, Reid, Marras, and Yi (2010). We plan to investigate these connections in future work.

Acknowledgement

The author thanks anonymous referees for comments that greatly improved the manuscript. We also thank Jessi Cisewski, Hari Iyer, Thomas Lee, Fuchang Gao and Marco Ferreira for helpful conversations and suggestions. Part of this work has been carried out while the author was visiting the Royal University of Phnom Penh, Cambodia. The author is thankful to Herb Clemens and the National Research Council for funding this trip. Jan Hannig's research was supported in part by the National Science Foundation under Grant No. 0707037 and 1007543.

Appendix A: Proof of Theorem 1.

The exact form of the generalized fiducial distribution (3.3) appears to be rather difficult to derive explicitly. Fortunately, one can find an explicit formula for the distribution of certain extremal points of the set $Q((\mathbf{a},\mathbf{b}),\mathbf{U}^\star)$.

We restrict our attention to $\mathbf{x} \in \mathcal{B}(\mathbf{x}_0)$. The assumptions guarantee that for every $\mathbf{u} \in \mathcal{U}_{\mathbf{i}}$ and $\mathbf{x}_{\mathbf{i}}$ the function

$$Q_{\mathbf{i}}(\mathbf{x}_{\mathbf{i}},\mathbf{u}) = \xi \quad \text{if} \quad \mathbf{G}_{\mathbf{i}}(\xi,\mathbf{u}) = \mathbf{x}_{\mathbf{i}}$$

is well defined. Moreover, for each fixed $\mathbf{u}$, the function $Q_{\mathbf{i}}(\,\cdot\,,\mathbf{u})$ is a homeomorphism and, for each fixed $\mathbf{x}_{\mathbf{i}}$, the function $Q_{\mathbf{i}}(\mathbf{x}_{\mathbf{i}},\,\cdot\,)$ is continuous. Moreover, for any $(\mathbf{a},\mathbf{b})$,

$$Q((\mathbf{a},\mathbf{b}),\mathbf{u}) = \bigcap_{\mathbf{i}} Q_{\mathbf{i}}((\mathbf{a}_{\mathbf{i}},\mathbf{b}_{\mathbf{i}}),\mathbf{u}).$$

Thus any point on the boundary of $Q((\mathbf{a},\mathbf{b}),\mathbf{u})$ is also on the boundary of $Q_{\mathbf{i}}((\mathbf{a}_{\mathbf{i}},\mathbf{b}_{\mathbf{i}}),\mathbf{u})$ for some $\mathbf{i}$. Let $C_{\mathbf{i}}$ denote the set of 2^p vertices of $(\mathbf{a}_{\mathbf{i}},\mathbf{b}_{\mathbf{i}})$, $C((\mathbf{a},\mathbf{b}),\mathbf{u}) = \bigcup_{\mathbf{i}} Q_{\mathbf{i}}(C_{\mathbf{i}},\mathbf{u})$ and $Q_E((\mathbf{a},\mathbf{b}),\mathbf{u}) = C((\mathbf{a},\mathbf{b}),\mathbf{u}) \cap Q([\mathbf{a},\mathbf{b}],\mathbf{u})$. Because of our assumptions on uniqueness of inverses, the set $Q_E((\mathbf{a},\mathbf{b}),\mathbf{u}) = \emptyset$ if and only if $Q((\mathbf{a},\mathbf{b}),\mathbf{u}) = \emptyset$. Moreover, the points in $Q_E((\mathbf{a},\mathbf{b}),\mathbf{u})$ lie on the boundary of $Q((\mathbf{a},\mathbf{b}),\mathbf{u})$. In fact, $Q_E((\mathbf{a},\mathbf{b}),\mathbf{u})$ could be viewed as the set of extremal points of $Q((\mathbf{a},\mathbf{b}),\mathbf{u})$.

Let $\mathbf{d} = \{\mathbf{d}_1,\ldots,\mathbf{d}_p\} \subset \mathbb{R}^p$ be a collection of orthonormal basis vectors. Take the furthest point in $Q_E((\mathbf{a},\mathbf{b}),\mathbf{u})$ along the direction $\mathbf{d}_1$. If there are ties, select among the tied points the one furthest along $\mathbf{d}_2$, etc. Eventually a unique value in $Q_E((\mathbf{a},\mathbf{b}),\mathbf{u})$ is selected. We denote it by $Q_{\mathbf{d}}((\mathbf{a},\mathbf{b}),\mathbf{u})$.

Similarly, for each $\mathbf{i}$ consider the furthest point in $Q_{\mathbf{i}}(C_{\mathbf{i}},\mathbf{u})$ along $\mathbf{d}$ and denote the vertex in $C_{\mathbf{i}}$ that maps to this extreme $\mathbf{c}_{\mathbf{i}}^{\mathbf{d}}(\mathbf{u})$.

Lemma A.1. *Under the assumptions of Theorem 1 the distribution of*

$$Q_{\mathbf{d}}((\mathbf{a},\mathbf{b}),\mathbf{U}^\star) \mid \{Q((\mathbf{a},\mathbf{b}),\mathbf{U}^\star) \neq \emptyset\} \tag{A.1}$$

is absolutely continuous with density

$$r_{\mathbf{d}}(\xi) \propto \sum_{\mathbf{i}} \int_{(\mathbf{a}_{\mathbf{i}^\complement},\mathbf{b}_{\mathbf{i}^\complement})} \sum_{\mathbf{s}_{\mathbf{i}} \in C_{\mathbf{i}}} f_{\xi \mathbf{X}_{\mathbf{i}^\complement}}(\xi,\mathbf{s}_{\mathbf{i}^\complement} \mid \mathbf{s}_{\mathbf{i}}) I_{\{\mathbf{c}_{\mathbf{i}}^{\mathbf{d}}(\mathbf{G}^{-1}((\mathbf{s}_{\mathbf{i}},\mathbf{s}_{\mathbf{i}^\complement}),\xi))=\mathbf{s}_{\mathbf{i}}\}} d\mathbf{s}_{\mathbf{i}^\complement},$$

where $f_{\xi \mathbf{X}_{\mathbf{i}^\complement}}(\xi,\mathbf{s}_{\mathbf{i}^\complement} \mid \mathbf{s}_{\mathbf{i}})$ *is given by* (3.2).

Proof. The conditional distribution (A.1) is well-defined, because the condition

$$P(Q((\mathbf{a},\mathbf{b}),\mathbf{U}^\star)\neq\emptyset)\geq P_{\xi_0}(\mathbf{X}\in(\mathbf{a},\mathbf{b}))>0.$$

The assumptions imply $Q_{\mathbf{d}}((\mathbf{a},\mathbf{b}),\mathbf{U}^\star)$ is equal to exactly one of the $\mathbf{c}_{\mathbf{i}}^{\mathbf{d}}(\mathbf{U}^\star)$ with probability 1.

Denote by $\mathbf{Y}_{\mathbf{i}}(\mathbf{u},\mathbf{s}_{\mathbf{i}})$ the unique solution $(\tilde{\xi},\tilde{\mathbf{s}}_{\mathbf{i}^\complement})$ to the equation $(\mathbf{s}_{\mathbf{i}},\tilde{\mathbf{s}}_{\mathbf{i}^\complement})=\mathbf{G}(\tilde{\xi},\mathbf{u})$. By assumptions, $\mathbf{Y}_{\mathbf{i}}(\mathbf{U}^\star,\mathbf{s}_{\mathbf{i}})$ is a random variable with density given by (3.2). Compute

$$\begin{aligned}
&P\Big(\{Q_{\mathbf{d}}((\mathbf{a},\mathbf{b}),\mathbf{U}^\star)\leq\mathbf{z}\}\cap\{Q((\mathbf{a},\mathbf{b}),\mathbf{U}^\star)\neq\emptyset\}\Big)\\
&\quad=\sum_{\mathbf{i}}\sum_{\mathbf{s}_{\mathbf{i}}\in C_{\mathbf{i}}}P(\{\mathbf{c}_{\mathbf{i}}^{\mathbf{d}}(\mathbf{U}^\star)=\mathbf{s}_{\mathbf{i}}\}\cap\{\mathbf{Y}_{\mathbf{i}}(\mathbf{U}^\star,\mathbf{s}_{\mathbf{i}})\in(-\infty,\mathbf{z})\times(\mathbf{a}_{\mathbf{i}^\complement},\mathbf{b}_{\mathbf{i}^\complement})\})\\
&\quad=\sum_{\mathbf{i}}\sum_{\mathbf{s}_{\mathbf{i}}\in C_{\mathbf{i}}}\int_{(-\infty,\mathbf{z})\times(\mathbf{a}_{\mathbf{i}^\complement},\mathbf{b}_{\mathbf{i}^\complement})}f_{\xi\mathbf{X}_{\mathbf{i}^\complement}}(\xi,\mathbf{s}_{\mathbf{i}^\complement}\mid\mathbf{s}_{\mathbf{i}})I_{\{\mathbf{c}_{\mathbf{i}}^{\mathbf{d}}(\mathbf{G}^{-1}((\mathbf{s}_{\mathbf{i}},\mathbf{s}_{\mathbf{i}^\complement}),\xi))=\mathbf{s}_{\mathbf{i}}\}}d\xi d\mathbf{s}_{\mathbf{i}^\complement}
\end{aligned}$$

The last step follows from (3.2) and the fact that for each $\mathbf{s}_{\mathbf{i}}$, there is a one-to-one map between $(\mathbf{s}_{\mathbf{i}^\complement},\xi)$ and $\mathbf{u}$. The proof now follows by differentiation.

Proof of Theorem 1. The assumptions of the theorem guarantee that $C_{\mathbf{i}}\to\{\mathbf{x}\}$. Thus $Q((\mathbf{a}_m,\mathbf{b}_m),U^\star)$, if non empty, converges to a singleton. To find the distribution of the limit it is enough to find the limiting distribution of

$$Q_{\mathbf{d}}((\mathbf{a}_m,\mathbf{b}_m),\mathbf{U}^\star)\mid\{Q((\mathbf{a}_m,\mathbf{b}_m),\mathbf{U}^\star)\neq\emptyset\}$$

for any fixed $\mathbf{d}$.

Fix ξ and recall that $\mathbf{x}_0$, the observed value of our data, is also fixed. The continuity of $f_{\xi\mathbf{X}_{\mathbf{i}^\complement}}(\xi,\mathbf{y}_{\mathbf{i}^\complement}|\mathbf{y}_{\mathbf{i}})$ implies that

$$\lim_{m\to\infty}\sup_{\mathbf{y}\in[\mathbf{a}_m,\mathbf{b}_m]}|f_{\xi\mathbf{X}_{\mathbf{i}^\complement}}(\xi,\mathbf{y}_{\mathbf{i}^\complement}|\mathbf{y}_{\mathbf{i}})-f_{\xi\mathbf{X}_{\mathbf{i}^\complement}}(\xi,\mathbf{x}_{0,\mathbf{i}^\complement}|\mathbf{x}_{0,\mathbf{i}})|=0,$$

and a simple calculation shows that for each $\mathbf{i}$,

$$\begin{aligned}
c_m^{n-p}\sum_{\mathbf{s}_{\mathbf{i}}\in C_i}\int_{(\mathbf{a}_{\mathbf{i}^\complement},\mathbf{b}_{\mathbf{i}^\complement})}f_{\xi\mathbf{X}_{\mathbf{i}^\complement}}(\xi,\mathbf{s}_{\mathbf{i}^\complement}\mid\mathbf{s}_{\mathbf{i}})I_{\{\mathbf{c}_{\mathbf{i}}^{\mathbf{d}}(\mathbf{G}^{-1}((\mathbf{s}_{\mathbf{i}},\mathbf{s}_{\mathbf{i}^\complement}),\xi))=\mathbf{s}_{\mathbf{i}}\}}\,d\mathbf{s}_{\mathbf{i}^\complement}&\\
\to f_{\xi\mathbf{X}_{\mathbf{i}^\complement}}(\xi,\mathbf{x}_{0,\mathbf{i}^\complement}\mid\mathbf{x}_{0,\mathbf{i}})\prod_{j\in\mathbf{i}^\complement}w_j.&
\end{aligned}$$

Similarly the assumption on the continuity of the integral implies

$$\begin{aligned}
c_m^{n-p}\int_\Xi\sum_{\mathbf{s}_{\mathbf{i}}\in C_i}\int_{(\mathbf{a}_{\mathbf{i}^\complement},\mathbf{b}_{\mathbf{i}^\complement})}f_{\xi\mathbf{X}_{\mathbf{i}^\complement}}(\xi,\mathbf{s}_{\mathbf{i}^\complement}\mid\mathbf{s}_{\mathbf{i}})I_{\{\mathbf{c}_{\mathbf{i}}^{\mathbf{d}}(\mathbf{G}^{-1}((\mathbf{s}_{\mathbf{i}},\mathbf{s}_{\mathbf{i}^\complement}),\xi))=\mathbf{s}_{\mathbf{i}}\}}\,d\mathbf{s}_{\mathbf{i}^\complement}\,d\xi&\\
\to\int_\Xi f_{\xi\mathbf{X}_{\mathbf{i}^\complement}}(\xi,\mathbf{x}_{0,\mathbf{i}^\complement}\mid\mathbf{x}_{0,\mathbf{i}})\prod_{j\in\mathbf{i}^\complement}w_j\,d\xi.&
\end{aligned}$$

The statement of the Theorem follows immediately.

Appendix B: Proof of Theorem 2

Let S_j, $j = 0, \ldots, k$, be the number of observations in $(a_j a_{j+1}]$. The distribution of $\mathbf{S}$ is multinomial$(n, p_0(\xi_0), \ldots, p_k(\xi_0))$. Just as in Appendix A, let $\mathbf{d} = \{\mathbf{d}_1, \ldots, \mathbf{d}_p\} \subset \mathbb{R}^p$ be a collection of orthonormal basis vectors. For $\mathbf{j} \subset \{1, \ldots, k\}$ and $\mathbf{t} \in \{0,1\}^p$, define $\mathbf{c}^{\mathbf{d}}_{\mathbf{j},\mathbf{t}}(\mathbf{u_j})$ as the the vertex in $(\mathbf{a_{i+t-1}}, \mathbf{a_{i+t}})$ that maps to the furthest point in $Q_{\mathbf{j}}(\{(\mathbf{a_{j+t-1}}, \mathbf{a_{j+t}}\}, \mathbf{u_j})$ along $\mathbf{d}$.

Lemma B.1. *Under the assumptions of Theorem* 2 *the distribution of* (A.1) *is absolutely continuous with density*

$$r_{\mathbf{d}}(\xi) \propto \frac{(2\pi/n)^{(k-p)/2}\Gamma(\sum_{i=0}^k S_i)}{\prod_{i=0}^k \Gamma(S_i)} \prod_{i=0}^k p_i(\xi)^{S_i-1}$$
$$\cdot \sum_{\mathbf{j}} J(\mathbf{a_j}, \xi)\Bigg(\sum_{\mathbf{t}\in\{0,1\}^p} I_{\{\mathbf{c}^{\mathbf{d}}_{\mathbf{j},\mathbf{t}}(\mathbf{G}^{-1}_{\mathbf{j}}(\mathbf{a_j},\xi))=\mathbf{a_j}\}} \prod_{j\in\mathbf{j}+\mathbf{t}-1} n^{-1}S_j \prod_{j\in\{0,\ldots,k\}\setminus\mathbf{j}+\mathbf{t}-1} p_j(\xi)\Bigg), \quad \text{(B.1)}$$

where the Jacobian

$$J((x_1, \ldots, x_p), \xi) = \left|\det \frac{\mathbf{d}(F(x_1,\xi), \ldots, F(x_p,\xi))}{\mathbf{d}\xi}\right|.$$

Proof. If $F(F^{-1}(\xi, u), \xi) = u$ for each fixed ξ, then the lemma follows immediately from a calculation analogous to the proof of Lemma A.1 by simply rearranging the non-zero terms and multiplying both numerator and denominator by a suitable constant.

Otherwise, we can find $\tilde{F}(s, \xi)$ so that $\tilde{F}(a_i, \xi) = F(a_,\xi)$ for all $i = 1, \ldots, k$ and $\tilde{F}(\tilde{F}^{-1}(\xi, u), \xi) = u$. This is achieved by redistributing jumps continuously over the intervals (a_i, a_{i+1}). Define $\tilde{X}_i = \tilde{F^{-1}}(U_i, \xi)$ and denote the corresponding inverse (2.1) by $\tilde{Q}$. For $\mathbf{a}, \mathbf{b} \in \{a_0, \ldots, a_{k+1}\}$, the inverse $Q((\mathbf{a}, \mathbf{b}), \mathbf{u}) = \tilde{Q}((\mathbf{a}, \mathbf{b}), \mathbf{u})$. Since we only observe $\mathbf{X} \in (\mathbf{a}, \mathbf{b})$, the generalized fiducial distributions (2.3) computed based on the structural equation for $\mathbf{X}$ and $\tilde{\mathbf{X}}$ are the same.

Proof Theorem 2. We prove the theorem in two steps. First, we prove Berstein-von Mises for some special points in $Q((\mathbf{a}, \mathbf{b}), u)$ and verify the conditions of Theorem 1 of Hannig (2009) for them. We only need to verify Assumptions 1 and 2, as Assumption 3, related to the shape of the confidence set, is assumed. Second, we show that the same is true for all the other points in $Q((\mathbf{a}, \mathbf{b}), u)$.

Take $\mathbf{p} = (p_0(\xi_0), \ldots, p_k(\xi_0))$ and Σ as the variance matrix of the Multinomial $(1, \mathbf{p})$ distribution. By the Skorokhod Representation Theorem we can find $\mathbf{S}$ Multinomial$(n, \mathbf{p})$ and $\mathbf{H}$ Normal$(0, \Sigma)$ such that $\mathbf{S} = n\mathbf{p} + n^{1/2}\mathbf{H} + o_{as}(n^{1/2})$,

$n \to \infty$. Recall that $S_0 = n - \sum_{j=1}^{k} S_j$, $p_0(\xi) = 1 - \sum_{j=1}^{k} p_j(\xi)$, and $H_0 = -\sum_{j=1}^{k} H_j$.

Let $\mathcal{R}_\xi^{\mathbf{d}}$ have the generalized fiducial distribution given by (B.1). The density of $n^{1/2}(\mathcal{R}_\xi^{\mathbf{d}} - \xi_0)$ is $g(\mathbf{z}) = r_{\mathbf{d}}(\xi_0 + n^{-1/2}\mathbf{z})n^{-p/2}$. We investigate the behavior of $g(\mathbf{z})$ as $n \to \infty$.

Set

$$g_{2,\mathbf{j}}(\xi) = J(\mathbf{a_j}, \xi)\Big(\sum_{\mathbf{t}\in\{0,1\}^p} I_{\{\mathbf{c}^{\mathbf{d}}_{\mathbf{j},\mathbf{t}}(\mathbf{G}_{\mathbf{j}}^{-1}((\mathbf{a_{j+t-1}},\mathbf{a_{j+t}}),\xi))=\mathbf{a_j}\}} \times \prod_{j\in\mathbf{j}+\mathbf{t}-1} n^{-1}S_j \prod_{j\in\{0,\ldots,n\}\setminus\mathbf{j}+\mathbf{t}-1} p_j(\xi)\Big).$$

By our assumptions, $g_{2,\mathbf{j}}(\xi_0 + n^{-1/2}\mathbf{z}) \to g_{2,\mathbf{j}}(\xi_0)$ a.s..

Compute, using Taylor series and Stirling's formula,

$$\begin{aligned}\log(g_1(\mathbf{z})) &= \log\left(\frac{n^{-p/2}(2\pi/n)^{(k-p)/2}\Gamma(\sum_{i=1}^{k} S_i)}{\prod_{i=1}^{k}\Gamma(S_i)}\prod_{i=0}^{k} p_i(\xi_0 + n^{-1/2}\mathbf{z})^{S_i-1}\right)\\ &= -\frac{p}{2}\log(2\pi) - \sum_{j=0}^{k} S_j\log(n^{-1}S_j) + \frac{1}{2}\sum_{j=0}^{k}\log(n^{-1}S_j)\\ &\quad + \sum_{j=0}^{k} S_j\log(p_j(\xi_0+n^{-1/2}\mathbf{z})) - \sum_{j=0}^{k}\log(p_j(\xi_0+n^{-1/2}\mathbf{z})) + o_{as}(1). \quad (B.2)\end{aligned}$$

By our assumptions

$$\frac{1}{2}\sum_{j=0}^{k}\log(n^{-1}S_j) - \sum_{j=0}^{k}\log(p_j(\xi_0 + n^{-1/2}\mathbf{z})) \to -\frac{1}{2}\sum_{j=0}^{k}\log(p_j(\xi_0)) \quad a.s..$$

Using $S_j = np_j(\xi_0) + n^{1/2}H_j + o_{as}(n^{1/2})$, we compute

$$\begin{aligned}&\sum_{j=0}^{k} S_j\log(n^{-1}S_j)\\ &= \sum_{j=0}^{k} S_j\left(\log(p_j(\xi_0)) + \frac{n^{-1}S_j - p_j(\xi_0)}{p_j(\xi_0)} - \frac{1}{2}\Big(\frac{n^{-1}S_j - p_j(\xi_0)}{p_j(\xi_0)}\Big)^2 + o_{as}(n^{-1})\right)\\ &= \sum_{j=0}^{k} S_j\log(p_j(\xi_0)) + \frac{1}{2}\sum_{j=0}^{k}\frac{H_j^2}{p_j(\xi_0)} + o_{as}(1).\end{aligned}$$

Using $p_j(\xi_0 + n^{-1/2}\mathbf{z}) = p_j(\xi_0) + n^{-1/2}\nabla p_j(\xi_0)\cdot\mathbf{z} + o(n^{-1/2})$, we analogously compute

$$\sum_{j=0}^{k} S_j \log(p_j(\xi_0 + n^{-1/2}\mathbf{z}))$$
$$= \sum_{j=0}^{k} S_j \log(p_j(\xi_0)) + \sum_{j=0}^{k} \frac{H_j(\nabla p_j(\xi_0)\cdot\mathbf{z})}{p_j(\xi_0)} - \frac{1}{2}\frac{(\nabla p_j(\xi_0)\cdot\mathbf{z})^2}{p_j(\xi_0)} + o_{as}(1).$$

By plugging back into (B.2) we get

$$g_1(\mathbf{z}) \to \frac{\exp\left(-\sum_{j=0}^{k}(\nabla p_j(\xi_0)\cdot\mathbf{z} - H_j)^2/2p_j(\xi_0)\right)}{(2\pi)^{p/2}\left(\prod_{j=0}^{k} p_j(\xi_0)\right)^{1/2}} \quad a.s.. \tag{B.3}$$

Denote the function on the right side of (B.3) by $\tilde{n}(\mathbf{z})$. We show this function is, up to a constant, a density of a non-degenerate, multivariate normal distribution.

The random vector $\tilde{\mathbf{H}} = (H_1, \ldots, H_k)$ is a non-degenarate Normal$(0, \tilde{\Sigma})$. Define the $p \times k$ Jacobi matrix

$$A = \left(\frac{\partial p_j(\xi_0)}{\partial \xi^r}\right)_{r=1,\ldots,p,\ j=1,\ldots,k},$$

the diagonal $k \times k$-matrix $D = \mathrm{diag}(p_1(\xi_0), \ldots, p_k(\xi_0))^{-1}$, and the $p \times p$-matrix $V = A\left(D + (1 - \sum_{j=1}^{k} p_j(\xi_0))^{-1}\mathbf{1}\cdot\mathbf{1}^\top\right)A^\top$. By our assumptions, A is full rank and V is non-singular, hence positive definite. A simple multiplication reveals that $\tilde{\Sigma}^{-1} = D + (1 - \sum_{j=1}^{k} p_j(\xi_0))^{-1}\mathbf{1}\cdot\mathbf{1}^\top$, so that $V = A\tilde{\Sigma}^{-1}A^\top$. Recall that properties of multinomial distribution imply $\det\tilde{\Sigma} = \prod_{i=0}^{k} p_i(\xi_0)$.

After some slightly tedious algebra we obtain that the function $n(\mathbf{z}) = C\tilde{n}(\mathbf{z})$,

$$C = \left(\det V \prod_{j=0}^{k} p_j(\xi_0)\right)^{1/2} \exp\left(\frac{1}{2}\tilde{\mathbf{H}}^\top\left(\tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1}A^\top V^{-1}A\tilde{\Sigma}^{-1}\right)\tilde{\mathbf{H}}\right),$$

is the density of a multivariate normal distribution with mean $V^{-1}A\tilde{\Sigma}^{-1}\tilde{\mathbf{H}}$ and variance matrix V^{-1}.

In particular, if $k = p$, $|\det A| = J((a_1, \ldots, a_p), \xi_0)$, and consequently $C = J((a_1, \ldots, a_p), \xi_0)$. Thus

$$g_1(\mathbf{z})J((a_1, \ldots, a_p), \xi_0 + n^{-3/2}\mathbf{z}) \to n(\mathbf{z}) \quad a.s.. \tag{B.4}$$

However, since the function $g_1(\mathbf{z})J((a_1, \ldots, a_p), \xi_0 + n^{-1\frac{1}{2}}\mathbf{z})$ is a transformation of Dirichlet density, it integrates to 1. Hence the convergence in (B.4) is also in L^1.

Since $0 \le n^{-1}S_j \le 1$ and $0 \le p_j(\xi) \le 1$, the uniform integrability of $g_1(\mathbf{z})g_{2,\mathbf{j}}(\mathbf{z})$ follows and one can conclude that $n^{1/2}(\mathcal{R}_\xi^{\mathbf{d}} - \xi_0)$ converges in distribution to a Normal with mean $V^{-1}A\tilde{\Sigma}^{-1}\tilde{\mathbf{H}}$ and variance matrix V^{-1} almost surely.

If $k > p$, we have for each $\mathbf{j}$, $g_1(\mathbf{z})g_{2,\mathbf{j}}(\mathbf{z}) \to D_{\mathbf{j}}\, n(\mathbf{z})$ *a.s.*, and

$$g_1(\mathbf{z})g_{2,\mathbf{j}}(\mathbf{z}) \le 2^k g_1(\mathbf{z})J(\mathbf{a_j}, \xi_0 + n^{-1/2}\mathbf{z}).$$

We now show that $g_1(\mathbf{z})J(\mathbf{a_j}, \xi_0 + n^{-1/2}\mathbf{z})$ is uniformly integrable by comparison with a density based on $k = p$. In order to do that, we pool the digitizing categories between the entries of $\mathbf{j}$, i.e.,

$$\tilde{p}_j(\xi) = p_{i_j}(\xi) + \cdots + p_{i_{j+1}-1}(\xi), \quad \tilde{S}_j = S_{i_j} + \cdots + S_{i_{j+1}-1},$$

where $i_0 = 0, i_{p+1} = k+1$, and

$$\tilde{g}_1(\mathbf{z}) = \frac{n^{-p/2}\Gamma(\sum_{i=1}^p \tilde{S}_i)}{\prod_{i=1}^p \Gamma(\tilde{S}_i)} \prod_{i=0}^p \tilde{p}_i(\xi_0 + n^{-1/2}\mathbf{z})^{S_i - 1}.$$

As argued above, $\tilde{g}_1(\mathbf{z})J(\mathbf{a_j}, \xi_0 + n^{-1/2}\mathbf{z})$ is uniformly integrable. Moreover, by the Stirling formula and simple algebra, there are constant K_1 and K_2 independent on $\mathbf{z}$ and n such that

$$\frac{g_1(\mathbf{z})}{\tilde{g}_1(\mathbf{z})} \le K_1 n^{-(k-p)/2} \frac{\prod_{i=1}^p \Gamma(\tilde{S}_i)}{\prod_{i=1}^k \Gamma(S_i)} \le K_2 \quad a.s..$$

Thus $g_1(\mathbf{z})g_{2,\mathbf{j}}(\mathbf{z})$ is uniformly integrable and $g_1(\mathbf{z})g_{2,\mathbf{j}}(\mathbf{z}) \to D_{\mathbf{j}}n(\mathbf{z})$ in L^1. We conclude by a straightforward calculation that $n^{1/2}(\mathcal{R}_\xi^{\mathbf{d}} - \xi_0)$ converges in distribution to a Normal with mean $V^{-1}A\tilde{\Sigma}^{-1}\tilde{\mathbf{H}}$ and variance matrix V^{-1} almost surely.

Finally, notice that $\mathrm{Var}\,(V^{-1}A\tilde{\Sigma}^{-1}\tilde{\mathbf{H}}) = V^{-1}$. The Assumptions 1, 2 of Theorem 1 in Hannig (2009) are verified for the special extreme points $\mathcal{R}_\xi^{\mathbf{d}}$. The first step of the proof is complete.

Now we finish the proof by showing that

$$\operatorname{diam} Q((\mathbf{a_k}, \mathbf{a_{k+1}}), \mathbf{U}^\star) \mid \{Q((\mathbf{a_k}, \mathbf{a_{k+1}}), \mathbf{U}^\star) \ne \emptyset\} = O_P(n^{-1}) \quad a.s.. \tag{B.5}$$

This, together with what was proved above, verifies the Assumptions 1, 2 of Theorem 1 in Hannig (2009) for $\mathcal{R}_\xi$ based on any point in $Q((\mathbf{a_k}, \mathbf{a_{k+1}}), \mathbf{U}^\star)$.

Recall that our observations are in the form $x_i \in (a_{k_i}, a_{k_i+1})$ for $i = 1, \ldots, n$. Notice that

$$\begin{aligned}
&P(\operatorname{diam} Q((\mathbf{a_k}, \mathbf{a_{k+1}}), \mathbf{U}^\star) > K/n \mid Q((\mathbf{a_k}, \mathbf{a_{k+1}}), \mathbf{U}^\star) \ne \emptyset) \\
&\quad = \int P(\operatorname{diam} Q((\mathbf{a_k}, \mathbf{a_{k+1}}), \mathbf{U}^\star) > K/n \mid Q_{\mathbf{d}}((\mathbf{a_k}, \mathbf{a_{k+1}}), \mathbf{U}^\star) = \xi) f_{\mathcal{R}^d}(\xi),
\end{aligned}$$

where $f_{\mathcal{R}^d}(\xi)$ is the density of $Q_{\mathbf{d}}((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)$ given $\{Q((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)\neq\emptyset\}$ displayed in (B.1). For $\mathbf{i}\subset\{1,\ldots,n\}$ denote by $J^{\mathbf{d}}_{\mathbf{i}}$ the event that the $Q_{\mathbf{d}}((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)$ is based of the p observational inequalities for $\mathbf{X_i}$. We then have

$$
\begin{aligned}
&P(\operatorname{diam} Q((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)>K/n\mid Q_{\mathbf{d}}((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)=\xi)\\
&\quad=\sum_{\mathbf{i}} P(\operatorname{diam} Q((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)>K/n\mid J^{\mathbf{d}}_{\mathbf{i}},\{Q_{\mathbf{d}}((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)=\xi\})\\
&\qquad\times P(J^{\mathbf{d}}_{\mathbf{i}}\mid Q_{\mathbf{d}}((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)=\xi).
\end{aligned}
$$

Consider $J^{\mathbf{d}}_{\mathbf{i}}$. The observational inequalities labeled by $a_{k_i}<x_i\le a_{k_i+1}$, $i\in\mathbf{i}$ are used for computing $Q_{\mathbf{d}}$. From the remaining observational inequalities we get $\mathbf{U}^\star_{\mathbf{i}^\complement}|J^{\mathbf{d}}_{\mathbf{i}}\cap\{Q_{\mathbf{d}}((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)=\xi\}$ are independent Uniform$(F(a_{k_j},\xi),F(a_{k_j+1},\xi))$ random variables, respectively, i.e., conditionally on $J^{\mathbf{d}}_{\mathbf{i}}\cap\{Q_{\mathbf{d}}((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)=\xi\}$, the random variables $\tilde U^\star_i=(U^\star_i-F(a_{k_i},\xi))/(F(a_{k_i+1},\xi)-F(a_{k_i},\xi))$, $i\in\mathbf{i}^\complement$, are i.i.d. Uniform$(0,1)$. For each $j=0,\ldots,k$, we have S_j observations in $(a_j,a_{j+1}]$. We lose at most one observation per group to be a part of $\mathbf{i}$. Consequently, on the set $J^{\mathbf{d}}_{\mathbf{i}}\cap\{Q_{\mathbf{d}}((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)=\xi\}$ we have

$$
\begin{aligned}
&Q((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)\\
&\subset\Bigg(\bigcap_{j=1}^{k}\{\bar\xi: F(a_j,\bar\xi)\le F(a_{k_j},\xi)+(F(a_{k_j+1},\xi)-F(a_{k_j},\xi))\tilde U^j_{1:S_j-1}\}\\
&\cap\bigcap_{j=0}^{k-1}\{\bar\xi: F(a_{j+1},\bar\xi)\ge F(a_{k_j+1},\xi)+(F(a_{k_j+1},\xi)-F(a_{k_j},\xi))(1-\tilde U^j_{S_j-1:S_j-1})\}\Bigg).
\end{aligned}
\tag{B.6}
$$

Here $\tilde U^j_{1:S_{j-1}}$ and $\tilde U^j_{S_{j-1}:S_{j-1}}$ are the order statistics of an array obtained by reordering $\tilde U^\star_i$, $\mathbf{i}\in\mathbf{i}^\complement$, so that they are grouped according to their observational inequality.

By (B.6) and the differential geometric structure of our manifolds around the true value ξ_0, there is an open neighborhood $\mathcal{N}$ of ξ_0 and a constant C such that, for all $\xi\in\mathcal{N}$ and all $\mathbf{i}$,

$$\operatorname{diam} Q((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)\le C\max\{\tilde U^j_{1:S_j-1},1-\tilde U^j_{S_j-1:S_j-1},\ j=0,\ldots,k\}.$$

This and the well known fact that $n\tilde U^j_{1:S_j-1}$ and $n(1-\tilde U^j_{S_j-1:S_j-1})$ converge in distribution to Exponential(1), imply that for every ϵ there is K, independent of $\mathbf{i}$, such that

$$P(\operatorname{diam} Q((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)>K/n\mid J^{\mathbf{d}}_{\mathbf{i}},Q_{\mathbf{d}}((\mathbf{a_k},\mathbf{a_{k+1}}),\mathbf{U}^\star)=\xi)\ge\varepsilon$$

for all n, $\mathbf{i}$ and $\xi \in \mathcal{N}$ a.s.. Here the a.s. is due to the fact that $S_j \to \infty$ only a.s..

However, as proved above, $Q_{\mathbf{d}}((\mathbf{a_k}, \mathbf{a_{k+1}}), \mathbf{U}^\star) \xrightarrow{P} \xi_0$ a.s., and (B.5) follows immediately. Here the *a.s.* comes from the assumption, $n^{-1/2}(\mathbf{S} - n\mathbf{p}) \to \mathbf{H}$ a.s., obtained from Skorokhod's representation.

References

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

Berger, J. O., Bernardo, J. M. and Sun, D. (2009). The formal definition of reference priors. *Ann. Statist.* **37**, 905-938.

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

Cordero, R., Seckmeyer, G. and Labbe, F. (2006). Effect of resolution on the uncertainty evaluation. *Metrologia* **43**, 33-38.

Davison, A. C., Fraser, D. A. S. and Reid, N. (2006). Improved likelihood inference for discrete data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68**, 495-508.

Dawid, A. P. and Stone, M. (1982). The functional-model basis of fiducial inference. *Ann. Statist.* **10**, 1054-1074.

Dempster, A. P. (1963). On direct probabilites. *J. Roy. Statist. Soc. Ser. B* **25**, 100-110.

Dempster, A. P. (2008). The Dempster-Shafer Calculus for Statisticians. *Internat. J. 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.

E, L., Hannig, J. and Iyer, H. K. (2009). Fiducial Generalized Confidence interval for Median Lethal Dose (LD50). preprint.

Efron, B. (1986). Why isn't everyone a Bayesian? *Amer. Statist.* **40**, 1-11. With discussion and a reply by the author.

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

Fraser, A. M., Fraser, D. A. S. and Staicu, A.-M. (2010). The second order ancillary: A differential view with continuity. *Bernoulli* **16**, 1208-1223.

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

Fraser, D. A. S., Reid, N., Marras, E. and Yi, G. Y. (2010). Default priors for Bayesian and frequentist inference. *j. Roy. Statist. Soc. Ser. B* **72**, 631-654.

Frenkel, R. B. and Kirkup, L. (2005). Monte Carlo-based estimation of uncertainty owing to limited resolution of digital instruments. *Metrologia* **42**, 27-30.

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

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

Hannig, J., Iyer, H. K. and Wang, J. C.-M. (2007). Fiducial approach to uncertainty assessment: accounting for error due to instrument resolution. *Metrologia* **44**, 476-483.

Hannig, J. and Lee, T. C. M. (2009). Generalized fiducial inference for wavelet regression. *Biometrika* **96**, 847-860.

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

Shafer, G. (2011). A betting interpretation of Dempster-Shafer degrees of belief. *Internat. J. Approximate Reasoning* **52**, 127-136.

Stein, C. M. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential methods in statistics*, Volume 16 of *Banach Center Publ.*, 485-514. Warsaw: Polish Scientific Publishers PWN.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604-608.

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.

Wandler, D. V. and Hannig, J. (2011). Fiducial inference on the largest mean of a multivariate normal distribution. *J. Multivariate Anal.* **102**, 87-104.

Wandler, D. V. and Hannig, J. (2012a). A fiducial approach to multiple comparisons. *J. Statist. Plann. Inference* **142**, 878-895.

Wandler, D. V. and Hannig, J. (2012b). Generalized fiducial confidence intervals for extremes. *Extremes* **15**, 67-87.

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

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

Willink, R. (2007). On the uncertainty of the mean of digitized measurements. *Metrologia* **44**, 73-81.

Witkowsky, V. and Wimmer, G. (2009). Interval estimation of the mean of a normal distribution based on quantized observations. *Mathematica Slovaca* **59**, 627-645.

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

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

Department of Statistics and Operation Research, CB#3260, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3260, USA.

E-mail: jan.hannig@unc.edu

(Received May 2011; accepted May 2012)