

# On Asymptotic Properties of Generalized Fiducial Inference for Discretized Data

Jan Hannig\*

*Department of Statistics and Operation Research, CB#3260  
The University of North Carolina at Chapel Hill  
Chapel Hill, NC 27599-3260  
e-mail: [jan.hannig@unc.edu](mailto:jan.hannig@unc.edu)  
<http://www.unc.edu/~hannig>*

**Abstract:** Most data of which a statistician comes into contact has been rounded off in some manner, i.e., we do not observe  $X = x$  but  $a < X \leq b$ . In this paper we investigate theoretical properties of the generalized fiducial distribution introduced in Hannig (2009) for such discretized data. Results are provided for both fixed sample size with increasing precision, and increasing sample size with fixed precision. In particular, we show that under very mild conditions the generalized fiducial distribution for i.i.d. discretized data will always lead to asymptotically correct inference. Resolution of several issues related to non-uniqueness of generalized fiducial distributions is also proposed.

**AMS 2000 subject classifications:** Primary 62A01; secondary 62F25.

**Keywords and phrases:** Generalized Fiducial Inference, Asymptotic Properties, Dempster-Shafer calculus.

## 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. Although he discussed fiducial inference in several subsequent papers, there appears to be no rigorous definition of a fiducial distribution for a vector parameter  $\theta$ . In the case of a one-parameter family of distributions, Fisher gave the following definition for a fiducial density  $f(\theta|x)$  of the parameter based on a single observation  $x$  for the case where the cdf  $F(x|\theta)$  is a monotonic decreasing function of  $\theta$ :

$$f(\theta|x) = -\frac{\partial F(x|\theta)}{\partial \theta}. \quad (1.1)$$

While Fisher provided more complex examples, until recently there was no formal definition of the fiducial distribution for a general model.

---

\*Jan Hannig's research was supported in part by the National Science Foundation under Grant No. 0707037.

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) attempted to develop a rigorous framework for making inferences along the lines of Fisher's fiducial inference. Fraser assumed that the statistical model was coupled with an additional group structure, e.g., location-scale model. Dawid and Stone (1982) provided further insight by studying situations where fiducial inference lead 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) discussed Dempster-Shafer calculus, which is closely related to fiducial inference. Even with these advancements, it is fair to say that fiducial inference failed to secure a place in mainstream statistics.

Tsui and Weerahandi (1989) proposed a new approach for constructing hypothesis tests using the concept of generalized  $P$ -values. This idea was later extended to a method of constructing generalized confidence intervals in Weerahandi (1993). 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 good theoretical and empirical properties (E *et al.*, 2008, 2009; Hannig and Lee, 2009; Wandler and Hannig, 2009). Finally, we remark that some of the results in this paper are also vaguely related to Denoeux (2006) who constructs Dempster-Shafer belief functions for multinomial proportions, though his aims and methods are different from ours.

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

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

where  $G(\cdot, \cdot)$  is termed structural equation, and  $\mathbf{U}$  is the random component of the structural equation whose distribution is completely known.

Suppose that the structural relation (1.2) can be inverted and the inverse  $G^{-1}(\cdot, \cdot)$  always exists. That is, for any observed  $\mathbf{x}$  and for any arbitrary  $\mathbf{u}$ ,  $\xi$  is obtained as  $\xi = G^{-1}(\mathbf{x}, \mathbf{u})$ . As the distribution of  $\mathbf{U}$  is completely known, one can always generate a random sample  $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_M$  from it. This random sample of  $\mathbf{U}$  is transformed into a random sample for  $\{\xi: \tilde{\xi}_1 = G^{-1}(\mathbf{x}, \tilde{\mathbf{u}}_1), \dots, \tilde{\xi}_M = G^{-1}(\mathbf{x}, \tilde{\mathbf{u}}_M)\}$ , which is called the fiducial sample. Estimates and confidence intervals for  $\xi$  can be obtained from this fiducial sample  $\tilde{\xi}_1, \dots, \tilde{\xi}_M$ .

However, the inverse  $G^{-1}(\cdot, \cdot)$  does not always exist. This can happen under two situations: for any particular  $\mathbf{u}$ , either there is more than one  $\xi$  satisfying (1.2), or there is no  $\xi$  satisfying (1.2).

The first situation can be dealt with either by picking one of the solutions or the use of Dempster-Shafer calculus (Dempster, 2008), see the end of Section 2

for more detail. Furthermore, for most parametric problems this non-uniqueness of inverse function disappears asymptotically, c.f., Section 5.

For the second situation, Hannig (2009) suggests removing the offending values of  $\mathbf{u}$  from the sample space and then re-normalizing the probabilities. Specifically, we generate  $\mathbf{u}$  conditional on the event that the inverse  $G^{-1}(\cdot, \cdot)$  exists. The rationale for this choice is that we know our data  $\mathbf{x}$  were generated with some  $\xi_0$  and  $\mathbf{u}_0$ , which implies there is at least one solution  $\xi_0$  satisfying (1.2) when the “true”  $\mathbf{u}_0$  is considered. Therefore, we restrict our attention to only those values of  $\mathbf{u}$  for which  $G^{-1}(\cdot, \cdot)$  exists. However, this set has in many practical situations probability zero, leading to non-uniqueness due to Borel paradox (Casella and Berger, 2002, Section 4.9.3). Borel paradox is the fact that when conditioning on an event of probability zero, one can obtain any answer.

Careful evaluation of the event “the inverse  $G^{-1}(\cdot, \cdot)$  exists” reveals that it has probability zero only if the probability of observing our data is zero. Hence Borel paradox will never be an issue for a generalized fiducial inference for discrete models. Moreover, any data that a statistician can come into contact with has been rounded off somehow, both by a measuring instrument and by storage on a computer. Mathematically speaking we do not observe  $X = x$  but  $a < X \leq b$ . Therefore, even for a continuous random variables, the probability of observing such discretized data is non-zero. Acknowledging the fact that all data is discretized in the definition of generalized fiducial distribution will remove the non-uniqueness due to the Borel paradox.

The main aim of this paper is to study general theoretical properties of the generalized fiducial distribution for discretized data. First, we study the limit of the generalized fiducial distribution for a fixed sample size of continuous random variables as the observational precision increases, i.e.,  $(b - a) \rightarrow 0$ . We then argue that the limiting distribution can be viewed as the generalized fiducial distribution for exactly observed data, hence resolving the Borel paradox. It is interesting to point out that the limit matches a suggestion on a particular, heuristically motivated, resolution of the Borel paradox suggested in Hannig (2009). Additionally, we show that the Bayesian posterior can be obtained as a generalized fiducial distribution provided we incorporate prior information into the model formulation.

Second, we study the asymptotics of the generalized fiducial distribution for i.i.d. data as the sample size goes to infinity while the precision with which we measure our data remains fixed. We show that under very mild conditions the generalized fiducial distribution will always lead 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 the generalized fiducial distribution.

There is a 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 for data generated data from the same distribution can lead to a different generalized fiducial distribution. While we do not resolve this issue in this paper, a partial resolution to this issue is presented in the case of i.i.d. observations.

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 our data increases. Section 4 describes how Bayesian inference can be obtained as a result of the generalized fiducial recipe. Finally, Section 5 explores large sample asymptotics for the generalized fiducial inference under the presence of discretized data. Section 6 concludes.

## 2. Generalized Fiducial Inference

Hannig (2009) proposes the following formal definition of the generalized fiducial recipe. 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. The equation (1.2) is termed the *structural equation*. We define for any measurable set  $A \in \mathbb{R}^n$  a set-valued function

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

The function  $Q(A, \mathbf{u})$  is the generalized inverse of the function  $G$ . Assume  $Q(\mathbf{x}, \mathbf{u})$  is a measurable function of  $u$ .

Assume that a data set was generated using (1.2) and it has been observed that the sample value of  $\mathbf{x} \in A$ . Clearly there the values of  $\xi$  and  $u$  used to generate the observed data will satisfy  $G(\xi, \mathbf{u}) \in A$ . This leads to the following definition of a generalized fiducial distribution for  $\xi$ :

$$Q(A, \mathbf{U}^*) \mid \{Q(A, \mathbf{U}^*) \neq \emptyset\}, \quad (2.2)$$

where  $\mathbf{U}^*$  is an independent copy of  $\mathbf{U}$ .

The following example provides a simple illustration of the definition of a generalized fiducial distribution.

*Example 2.1.* Suppose  $\mathbf{U} = (U_1, \dots, U_n)$  where  $U_i$  are i.i.d.  $N(0, 1)$  and

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

for some  $\mu \in \mathbb{R}$  so the  $X_i$ s are i.i.d.  $N(\mu, 1)$ . We observe a discretized realization of  $\mathbf{X}$ , i.e., the only thing we know about the realized value  $\mathbf{x}$  is that  $x_i \in (a_i, b_i)$  for all  $i = 1, \dots, n$ . Set  $\mathbf{a} = (a_1, \dots, a_n)$  and  $\mathbf{b} = (b_1, \dots, b_n)$ .

If  $n = 1$  then  $Q((a, b), u) = (a - u, b - u)$  and  $P(Q((a, b), U^*) \neq \emptyset) = 1$ . Thus following (2.2) the generalized fiducial distribution is the distribution of the random interval  $(a - U^*, b - U^*)$ , where  $U^* \sim N(0, 1)$  independent of the data.

If  $n > 1$ , the define  $L(\mathbf{a}, \mathbf{u}) = \max_i \{a_i - u_i\}$  and  $R(\mathbf{b}, \mathbf{u}) = \min_j \{b_j - u_j\}$ . Then the set value inverse

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

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

$$\begin{aligned} P(L(\mathbf{a}, \mathbf{U}^*) \leq l, r < R(\mathbf{b}, \mathbf{U}^*)) &= P(a_i - l \leq U_i^* < b_i - r, \text{ for all } i) \\ &= \prod_{i=1}^n (\Phi(b_i - r) - \Phi(a_i - l))_+ \end{aligned} \quad (2.3)$$

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

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

Notice that if  $Q((\mathbf{a}, \mathbf{b}), \mathbf{u})$  is non-empty, it is an entire interval. There are two ways to interpret and use the generalized fiducial distribution. Either one can simply select a particular value inside the set  $Q(A, \mathbf{u})$  according to some, possibly random rule. While a choice of one point in  $Q(A, \mathbf{u})$  over another is arbitrary it will be shown that its influence diminishes asymptotically.

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

give a rule on how to interpret this measure. While Dempster-Shafer approach is theoretically appealing in its honest acknowledgment of the fact that there is some chance that we cannot say whether statement is true or not, we choose, in most practical applications, to resolve this uncertainty by simply selecting a particular value inside the set  $Q(A, u)$ .

### 3. Increasing Precision Asymptotics

In this section we discuss the behavior of the generalized fiducial distribution as we increase the precision of the measurements. Such an asymptotics is not very relevant for discrete distributions and therefore we will turn our attention to continuous distributions.

Let us suppose that the parameter of interest  $\xi \in \Xi \subset \mathbb{R}^p$  is  $p$ -dimensional. Recall the structural equation (1.2), and assume that  $\mathbf{U} \in \mathbb{R}^n$  is a continuous random vector with joint density  $f_{\mathbf{U}}(\mathbf{u})$ , which is continuous on its support. Write  $\mathbf{G} = (g_1, \dots, g_n)$  so that  $X_i = g_i(\xi, \mathbf{U})$  for  $i = 1, \dots, n$ .

Assume that for each fixed  $\xi \in \Xi$  the function  $\mathbf{G}(\xi, \cdot)$  is one-to-one and continuously differentiable. Thus, using the Jacobian transformation, the density of  $\mathbf{X}$  is

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

Assume that for all  $p$ -tuples of indexes  $\mathbf{i} = (i_1 < \dots < i_p) \subset \{1, \dots, n\}$  and each fixed  $\mathbf{x}_{\mathbf{i}} = (x_{i_1}, \dots, x_{i_p})$  the function  $\mathbf{G}^{-1}((\mathbf{x}_{\mathbf{i}}, \cdot), \cdot)$ , viewed as a function of  $\xi$  and  $\mathbf{x}_{\mathbf{i}^c}$ , is one-to-one and differentiable. Thus, the density of  $(\xi, \mathbf{X}_{\mathbf{i}^c})$  is

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

It follows that for any fixed  $\xi$  the function  $f_{\xi \mathbf{X}_{\mathbf{i}^c}}(\xi, \mathbf{x}_{\mathbf{i}^c} | \mathbf{x}_{\mathbf{i}})$  is continuous in  $\mathbf{x} = (\mathbf{x}_{\mathbf{i}}, \mathbf{x}_{\mathbf{i}^c})$ . Assume additionally that the marginal density  $\int_{\Xi} f_{\xi \mathbf{X}_{\mathbf{i}^c}}(\xi, \mathbf{x}_{\mathbf{i}^c} | \mathbf{x}_{\mathbf{i}}) d\xi$  is continuous in  $\mathbf{x} = (\mathbf{x}_{\mathbf{i}}, \mathbf{x}_{\mathbf{i}^c})$ .

We observe a discretized realization of  $\mathbf{X}$  generated using some true unknown  $\xi_0$ . In particular, the only knowledge we have about the realized value  $\mathbf{x}$  is that  $x_i \in (a_i, b_i)$  for all  $i = 1, \dots, n$ . Set  $\mathbf{a} = (a_1, \dots, a_n)$  and  $\mathbf{b} = (b_1, \dots, b_n)$ . We assume that  $P_{\xi_0}(\mathbf{X} \in (\mathbf{a}, \mathbf{b})) > 0$  which causes the conditional distribution in (2.2) to be uniquely defined.

**Theorem 3.1.** *Suppose all the assumptions stated in this section. Additionally, consider a sequence of  $p$ -dimensional intervals  $(\mathbf{a}_1, \mathbf{b}_1) \supset (\mathbf{a}_2, \mathbf{b}_2) \supset \dots$  and numbers  $c_m \uparrow \infty$  such that  $\bigcap_m (\mathbf{a}_m, \mathbf{b}_m) = \{\mathbf{x}\}$  and  $c_m(b_{m,i} - a_{m,i}) \rightarrow w_i > 0$  for all  $i = 1, \dots, m$ . Then the generalized fiducial distribution*

$$Q((\mathbf{a}_m, \mathbf{b}_m), U^*) | \{Q((\mathbf{a}_m, \mathbf{b}_m), U^*) \neq \emptyset\} \quad (3.3)$$

*converges to a singleton that has a continuous distribution with density*

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

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

$$J(\mathbf{x}, \xi) = \sum_{\mathbf{i}=(i_1, \dots, i_p)} \frac{1}{w_{i_1} \dots w_{i_p}} \left| \frac{\det \left( \frac{d}{d(\xi, \mathbf{x}_i \mathbf{0})} \mathbf{G}^{-1}(\mathbf{x}_i, \xi) \right)}{\det \left( \frac{d}{d\mathbf{x}} \mathbf{G}^{-1}(\mathbf{x}_i, \xi) \right)} \right|. \quad (3.5)$$

As discussed above, the definition of the generalized fiducial distribution (2.2) suffers non-uniqueness due to the Borel paradox if  $P(Q(A, U^*) \neq \emptyset) = 0$ . The Borel paradox states that there could be multiple valid interpretation of the conditional probability each leading to a different answer.

When all the observed data is on the same scale, such as in the case of i.i.d. observations, it is natural to require that  $w_1 = \dots = w_n = 1$ . Theorem 3.1 then provides a very natural unique way to define the conditional probability in (2.2) as (3.3). We remark that this agrees with and validates a heuristically motivated proposal of Hannig (2009). Using this suggestion E *et al.* (2008), Hannig and Lee (2009), and others used generalized fiducial distribution to propose confidence intervals in a wide variety of statistical problems.

As discussed in Hannig (2009), there are three sources of non-uniqueness in the definition of the fiducial distribution: the choice of structural equation, the Borel paradox if  $P(Q(A, U^*) \neq \emptyset) = 0$ , and the choice of a particular value in  $Q(A, U^*)$  if it contains more than one element. Theorem 3.1 gives a reasonable, consistent way of resolving the non-uniqueness due to the last two issues for a large class of problems.

In many practical applications the physical process by which the data was generated is known, hence the choice of the structural equation should reflect this process eliminating the problem of non-uniqueness due to the choice of structural equation. A canonical example arises in the field of metrology where an unknown quantity is measured using some known processes. The processes have known physical characteristics and add errors to the measured quantity in some pre-specified known fashion. The resulting measured values are expressed as an equation combining some unknown measured quantities and errors. This equation can be taken as the structural equation (1.2).

If we know the distribution of the data, and the realizations are i.i.d., we can follow Fisher's original definition. In this case we use the generalized inverse of the distribution function  $F^{-1}(\xi, u)$  to define the structural equation

$$X_i = F^{-1}(\xi, U_i) \quad \text{for } i = 1, \dots, n, \quad (3.6)$$

If additionally, the assumptions of Theorem 3.1 are satisfied, the generalized fiducial distribution will be (3.4) with (3.5) simplified to

$$J(\mathbf{x}, \xi) = \sum_{\mathbf{i}=(i_1, \dots, i_p)} \left| \frac{\det \left( \frac{d}{d\xi} (F(x_{i_1}, \xi), \dots, F(x_{i_p}, \xi)) \right)}{f(x_{i_1}, \xi) \dots f(x_{i_p}, \xi)} \right|. \quad (3.7)$$

Here  $F(x, \xi)$  and  $f(x, \xi)$  stand for the distribution and density functions of our data respectively. If  $n = p = 1$  (3.4) and (3.7) simplify to (1.1). Based on results in Section 5, this choice of the structural equation has favorable theoretical properties and we recommend it as the default.

#### 4. Bayesian posterior and Generalized Fiducial Distribution

Hannig *et al.* (2009) show that a fiducial predictive distribution for future or unobserved data can be easily obtained from the generalized fiducial recipe. This is achieved by assuming that the data in (1.2) is  $\mathbf{X} = (\mathbf{X}^o, \mathbf{X}^p)$ , where  $\mathbf{X}^o$  is observed and  $\mathbf{X}^p$  is unobserved and needs to be predicted. The structural equation becomes

$$\mathbf{X}^o = \mathbf{G}^o(\xi, U), \quad \mathbf{X}^p = \mathbf{G}^p(\xi, U), \quad (4.1)$$

where  $(\xi, \mathbf{X}^p)$  are the unknown parameters and  $U$  has fully known distribution. After observing  $\mathbf{x}^o \in A$ , for some measurable set  $A$ , the set inverse (2.1) becomes

$$Q(A, u) = \{(\mathbf{x}^p, \xi) : \mathbf{G}^o(\xi, u) \in A, \mathbf{G}^p(\xi, u) = \mathbf{x}^p\} \quad (4.2)$$

and the joint fiducial distribution of the parameters and the predicted values is given by

$$Q(A, U^*) \mid \{Q(A, U^*) \neq \emptyset\}. \quad (4.3)$$

The predicted fiducial distribution of  $\mathbf{X}_0$  is the marginal conditional distribution of (4.3).

In order to incorporate Bayesian assumption into this paradigm, we use the following structural equation.

$$\xi = \mathbf{E}, \quad \mathbf{X} = \mathbf{G}^o(\xi, \mathbf{U}). \quad (4.4)$$

Here  $(E, U)$  is a random vector with a fully known distribution. Without loss of generality assume that  $E$  has density  $\pi(\mathbf{e})$  with respect to a  $\sigma$ -finite measure  $\lambda$ .

This is in line with Bayesian belief that the unobserved parameter value  $\xi$  was generated by a random variable  $E$  with a fully known distribution and needs to be predicted based on the observed data. We will show that the predictive fiducial distribution of  $\xi$  is exactly the Bayesian posterior.

Notice that for any measurable  $A$

$$Q(A, (\mathbf{e}, \mathbf{u})) = \{\xi : \xi = \mathbf{e}, \mathbf{G}^o(\xi, \mathbf{u}) \in A\},$$

where  $Q(A, (\mathbf{e}, \mathbf{u}))$  is either a singleton or an empty set.

We observe a discretized realization of  $\mathbf{X}$  generated using (4.4). In particular, the only thing we know about the realized value  $\mathbf{x}$  is that  $x_i \in (a_i, b_i)$  for all  $i = 1, \dots, n$ . Set  $\mathbf{a} = (a_1, \dots, a_n)$  and  $\mathbf{b} = (b_1, \dots, b_n)$ . Denote by  $P(\mathbf{X} \in (\mathbf{a}, \mathbf{b}) \mid \xi)$  the probability with  $\mathbf{X}$  generated using a fixed value  $\xi$  and assume that  $P(\mathbf{X} \in (\mathbf{a}, \mathbf{b})) = \int_{\Xi} P(\mathbf{X} \in (\mathbf{a}, \mathbf{b}) \mid \xi) \pi(d\xi) > 0$ . This rules out some pathological situations and causes the conditional distribution in (4.3) to be uniquely defined. A simple calculation shows that the conditional distribution in (4.3) has a density with respect to  $\lambda$

$$r(\xi) = \frac{P(\mathbf{X} \in (\mathbf{a}, \mathbf{b}) \mid \xi) \pi(d\xi)}{\int_{\Xi} P(\mathbf{X} \in (\mathbf{a}, \mathbf{b}) \mid \xi') \pi(d\xi')}.$$

Finally, assume that for each fixed  $\xi$ ,  $\mathbf{X}$  has a density  $f(\mathbf{x} \mid \xi)$  with respect to some  $\sigma$ -finite measure  $\mu$ . We have the following theorem.



**Theorem 4.1.** *Suppose all the assumptions stated in this section. Additionally, consider a sequence of  $p$ -dimensional intervals  $(\mathbf{a}_1, \mathbf{b}_1) \supset (\mathbf{a}_2, \mathbf{b}_2) \supset \dots$  and numbers  $c_m \uparrow \infty$  such that  $\bigcap_m (\mathbf{a}_m, \mathbf{b}_m) = \{\mathbf{x}\}$  and  $c_m(b_{m,i} - a_{m,i}) \rightarrow w_i > 0$  for all  $i = 1, \dots, m$ . Then the generalized fiducial distribution*

$$Q((\mathbf{a}_m, \mathbf{b}_m), U^*) \mid \{Q((\mathbf{a}_m, \mathbf{b}_m), U^*) \neq \emptyset\} \quad (4.5)$$

converges to a distribution with density

$$\pi(\xi \mid \mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x} \mid \xi) \pi(\xi)}{\int_{\Xi} f_{\mathbf{X}}(\mathbf{x} \mid \xi') \pi(\xi') d\xi'} \quad (4.6)$$

with respect to the measure  $\lambda$  for  $\mu$ -almost all  $\mathbf{x}$ .

*Proof.* Radon-Nikodyms theory implies that  $\mu(\mathbf{a}_m, \mathbf{b}_m)^{-1} P(\mathbf{X} \in (\mathbf{a}, \mathbf{b}) \mid \xi) \rightarrow f_{\mathbf{X}}(\mathbf{x} \mid \xi)$ ,  $\mu$ -a.e. Similarly, Fubini's theorem and Radon-Nikodym's theory imply  $\mu(\mathbf{a}_m, \mathbf{b}_m)^{-1} \int_{\Xi} P(\mathbf{X} \in (\mathbf{a}, \mathbf{b}) \mid \xi') \pi(d\xi') \rightarrow \int_{\Xi} f_{\mathbf{X}}(\mathbf{x} \mid \xi') \pi(\xi') d\xi'$ ,  $\mu$ -a.e. and the theorem follows.  $\square$

Equation (4.6) is exactly the Bayesian posterior and we have received a Bayesian posterior as a result of the generalized fiducial argument. The model shown in (4.4) can be used if we have prior information on some of the parameters and not on others. A precursor to this approach can be found in Wang and Iyer (2006). We will demonstrate this on the following simple example.

*Example 4.1.* Let  $X_i$  be i.i.d.  $N(\mu, \sigma^2)$ . Additionally assume that there is a prior information on  $\mu \sim \pi(\mu)$  but no prior information on  $\sigma$ . We set up the following structural equation

$$X_i = \mu + \sigma Z_i, \quad \mu = M,$$

where  $Z_i$  are i.i.d.  $N(0,1)$  and  $M$  is a random variable with a density  $\pi(m)$ . Assume that the value of  $\mathbf{x}$  is observed and  $f(\mathbf{x} \mid \mu, \sigma)$  is the  $N(\mu, \sigma^2)$  density. Following the derivations in Section 3 and above we get that the generalized fiducial distribution of  $(\mu, \sigma)$  is

$$r(\mu, \sigma) = \frac{f(\mathbf{x} \mid \mu, \sigma) \sigma^{-1} J(\mu) \pi(\mu)}{\int_{-\infty}^{\infty} \int_0^{\infty} f(\mathbf{x} \mid \mu', \sigma') \sigma'^{-1} J(\mu') \pi(\mu') d\sigma' d\mu'}, \quad (4.7)$$

where  $J(\mu) = \sum_{i=1}^n |x_i - \mu| + \sum_{1 \leq i < j \leq n} |x_i - x_j|$ . The predictive fiducial distribution of  $\mu$  is obtained by integrating (4.7). In particular it is proportional to

$$r(\mu) \propto \frac{J(\mu) \pi(\mu)}{(\sum_{i=1}^n (x_i - \mu)^2)^{n/2}}.$$

This is not a Bayesian posterior with respect to any prior.

Recall that the generalized fiducial distribution with both  $\mu$  and  $\sigma$  unknown gives

$$r(\mu) \propto \frac{1}{(\sum_{i=1}^n (x_i - \mu)^2)^{n/2}}$$

which leads to the usual  $t_{n-1}$  inference.

We performed a limited simulation study to determine the repeated sample frequentist performance of the predictive fiducial distribution for  $\mu$ . In contrast to the usual good performance of generalized fiducial distribution, the predictive fiducial distribution for  $\mu$  appears to be somewhat conservative for highly informative priors and small sample sizes. More research into the interface between generalized fiducial and Bayesian inference is needed.

## 5. Increasing Sample Size Asymptotics

In this section we will 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. A special case of this situation has been studied in [Hannig et al. \(2007\)](#) where a thorough simulation study for discretized normal data was reported. It shows very good repeated sample frequentist performance for confidence intervals based on the generalized fiducial distribution. [Theorem 5.1](#) provides theoretical justification for their findings.

Let us assume the structural equation [\(3.6\)](#), i.e.,

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

where  $X_i$  are random variables,  $\xi \in \Xi \subset \mathbb{R}^p$  is a  $p$ -dimensional parameter and  $U_i$  are i.i.d.  $U(0, 1)$ . 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 would be required.

Assume  $F(x, \xi)$  is continuously differentiable in  $\xi$  for all  $x$ , i.e., all  $p$  partial derivatives are continuous. For each  $(x_1, \dots, x_p)$  the map  $(F(x_1, \xi), \dots, F(x_p, \xi)) = (u_1, \dots, u_p)$ , taken as a function of  $\xi$  is one to one.

Let us assume that  $\mathbb{R}$  is partitioned into a fixed grid

$$(-\infty, a_1], (a_1, a_2], \dots, (a_k, \infty).$$

Set  $a_0 = -\infty$ ,  $a_{k+1} = \infty$ . Assume that  $p_j(\xi) = P(X \in (a_j, a_{j+1}]) > 0$  for all  $j = 0, \dots, k$  and all  $\xi$ . Additionally, assume for all  $\mathbf{j} = (j_1 < \dots < j_p) \subset \{1, \dots, k\}$ , the  $p \times p$  Jacobian

$$\det \left( \frac{\mathbf{d}F(\mathbf{a}_{\mathbf{j}}, \xi_0)}{\mathbf{d}\xi} \right) \neq 0.$$

The values of  $X_i$  are observed only up to the resolution of the grid, i.e., we do not observe the realized value  $x_i$  itself, only which of the intervals it falls into, i.e.,  $x_i \in (a_{k_i}, a_{k_i+1}]$  for some  $(k_1, \dots, k_n)$ .

Finally, let  $\mathcal{R}_\xi$  be a random variable having the fiducial distribution

$$V(Q((\mathbf{a}_{\mathbf{k}}, \mathbf{a}_{\mathbf{k}+1}], U^*)) \mid \{Q((\mathbf{a}_{\mathbf{k}}, \mathbf{a}_{\mathbf{k}+1}], U^*) \neq \emptyset\},$$

where  $V(A)$  is a possibly random rule selecting an element from a set  $A$ , c.f. [Hannig \(2009\)](#).

**Theorem 5.1.** *Suppose all the assumptions in this section. Then any confidence set of a shape satisfying Assumption 3 of Hannig (2009) based on  $\mathcal{R}_\xi$  will have asymptotically correct coverage.*

The proof of the theorem is a generalization of a proof found in E *et al.* (2009) and is relegated to Appendix B. While the proof is quite lengthy, its main ideas can be summarized as follows. Let  $S_j$ ,  $j = 0, \dots, k$ , denote the number of observations in  $(a_j a_{j+1}]$ . Assume without loss of generality that  $n^{-\frac{1}{2}}(\mathbf{S} - n\mathbf{p}) \rightarrow \mathbf{H}$  a.s. Since  $n^{-\frac{1}{2}}(\mathbf{S} - n\mathbf{p}) \xrightarrow{\mathcal{D}} \mathbf{H}$ , the a.s. convergence is achieved by Skorokhod's representation theorem without affecting the coverage of confidence intervals which only depends on the distribution of  $\mathbf{S}$ . The proof is done in two steps. First we show that  $n^{\frac{1}{2}}(\mathcal{R}_\xi - \xi_0)$  is asymptotically normal a.s. and leads to correct confidence intervals for a particular choice of  $V(\cdot)$ . Then we show that  $n \text{diam } Q((\mathbf{a}_k, \mathbf{a}_{k+1}], U^*) \mid \{Q((\mathbf{a}_k, \mathbf{a}_{k+1}], U^*) \neq \emptyset\})$  is bounded in probability a.s., implying that  $n^{\frac{1}{2}}(\mathcal{R}_\xi - \xi_0)$  has the same limit regardless of choice of  $V(\cdot)$ .

We also remark that, as a consequence of this result, the do-not-know probability in Dempster-Shafer calculus (Dempster, 2008) will vanish and have no influence on setting generalized fiducial confidence intervals.

## 6. Conclusions

In this paper we have studied asymptotic properties of generalized fiducial distribution for discretely observed data. Use of discretized data is natural, because all data has been discretized due to instrument precision and computer storage.

We showed that acknowledging this fact allowed us to resolve several non-uniqueness problems in the definition of generalized fiducial distribution. We also showed that under some mild conditions, the generalized fiducial distribution for parametric problems and discretized data will always lead 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 *et al.* (2007) implement a generalization of a Gibb's sampler for discretely observed normal data and show that generalized fiducial inference for fat 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 (2009) and elsewhere.

## Acknowledgements

The author thanks Jessi Cisewski, Hari Iyer, Thomas Lee and Fuchang Gao 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.

## Appendix A: Proof of Theorem 3.1

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

The assumptions guarantee that for every  $\mathbf{u}$  and  $\mathbf{x}_i$  the function

$$Q_i(\mathbf{x}_i, \mathbf{u}) = \xi \quad \text{if} \quad \mathbf{G}_i(\xi, \mathbf{u}) = \mathbf{x}_i$$

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

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

Thus any point on the boundary of  $Q((\mathbf{a}, \mathbf{b}), \mathbf{u})$  is also on the boundary of  $Q_i((\mathbf{a}_i, \mathbf{b}_i), \mathbf{u})$  for some  $\mathbf{i}$ . Let  $C_i$  denote the set of  $2^p$  vertices of  $(\mathbf{a}_i, \mathbf{b}_i)$ ,  $C((\mathbf{a}, \mathbf{b}), \mathbf{u}) = \bigcup_{\mathbf{i}} Q_i(C_i, \mathbf{u})$  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  $\mathbf{d} = \{\mathbf{d}_1, \dots, \mathbf{d}_p\} \subset \mathbb{R}^p$  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  $Q_E((\mathbf{a}, \mathbf{b}), \mathbf{u})$  will be selected. We denote it by  $Q_{\mathbf{d}}((\mathbf{a}, \mathbf{b}), \mathbf{u})$ .

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

**Lemma A.1.** *Under the assumptions of Theorem 3.1 the distribution of*

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

*is continuous with density proportional to*

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

where  $f_{\xi \mathbf{X}_i}(\xi, \mathbf{s}_i \mid \mathbf{c}_i^{\mathbf{d}})$  is given by (3.2).

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

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

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

Denote  $\mathbf{Y}_i(\mathbf{u}, \mathbf{s}_i)$  the unique solution to  $(\mathbf{s}_i, \cdot) = \mathbf{G}(\cdot, \mathbf{u})$ . By assumptions,  $\mathbf{Y}_i(\mathbf{U}^*, \mathbf{s}_i)$  is a random variable with density given by (3.2). Compute

$$\begin{aligned} & P\left(\{Q_{\mathbf{d}}((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) \leq \mathbf{z}\} \cap \{Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) \neq \emptyset\}\right) \\ &= \sum_{\mathbf{i}} \sum_{\mathbf{s}_i \in C_i} P(\{\mathbf{c}_i^{\mathbf{d}}(\mathbf{U}^*) = \mathbf{s}_i\} \cap \{\mathbf{Y}_i(\mathbf{U}^*, \mathbf{s}_i) \in (-\infty, \mathbf{z}) \times (\mathbf{a}_{i\mathbb{C}}, \mathbf{b}_{i\mathbb{C}})\}) \\ &= \sum_{\mathbf{i}} \sum_{\mathbf{s}_i \in C_i} \int_{(-\infty, \mathbf{z}) \times (\mathbf{a}_{i\mathbb{C}}, \mathbf{b}_{i\mathbb{C}})} f_{\xi \mathbf{X}_{i\mathbb{C}}}(\xi, \mathbf{s}_{i\mathbb{C}} \mid \mathbf{s}_i) I_{\{\mathbf{c}_i^{\mathbf{d}}(\mathbf{G}^{-1}((\mathbf{s}_i, \mathbf{s}_{i\mathbb{C}}), \xi)) = \mathbf{s}_i\}} d\xi d\mathbf{s}_{i\mathbb{C}} \end{aligned}$$

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

*Proof of Theorem 3.1.* The assumptions of the theorem guarantee that  $C_i \rightarrow \{\mathbf{x}\}$ . Thus the fact that  $Q((\mathbf{a}_m, \mathbf{b}_m), \mathbf{U}^*)$ , if non empty, converges to a singleton follows from the assumptions. Thus to find the distribution of the limit it is enough to find the limiting distribution of

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

for any fixed  $\mathbf{d}$ .

Fix  $\xi$  and recall that  $\mathbf{x}$ , the observed value of our data, is also fixed. The continuity of  $f_{\xi \mathbf{X}_{i\mathbb{C}}}(\xi, \mathbf{y}_{i\mathbb{C}} \mid \mathbf{y}_i)$  implies that

$$\lim_{m \rightarrow \infty} \sup_{\mathbf{y} \in [\mathbf{a}_m, \mathbf{a}_n]} |f_{\xi \mathbf{X}_{i\mathbb{C}}}(\xi, \mathbf{y}_{i\mathbb{C}} \mid \mathbf{y}_i) - f_{\xi \mathbf{X}_{i\mathbb{C}}}(\xi, \mathbf{x}_{i\mathbb{C}} \mid \mathbf{x}_i)| = 0$$

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

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

Similarly the assumption on the continuity of the integral implies

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

and the statement of the Theorem follows immediately.  $\square$

## Appendix B: Proof of Theorem 5.1

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

**Lemma B.1.** *Under the assumptions of Theorem 5.1 the distribution of (A.1) is continuous with density proportional to*

$$\begin{aligned} r_{\mathbf{d}}(\xi) \propto & \frac{(2\pi/n)^{\frac{k-p}{2}} \Gamma(\sum_{i=1}^k S_i)}{\prod_{i=1}^k \Gamma(S_i)} \prod_{i=0}^k p_i(\xi)^{S_i-1} \\ & \times \sum_{\mathbf{j}} J(\mathbf{a}_{\mathbf{j}}, \xi) \left( \sum_{\mathbf{t} \in \{0,1\}^p} I_{\{\mathbf{c}_{\mathbf{j}, \mathbf{t}}^{\mathbf{d}}(\mathbf{G}_j^{-1}(\mathbf{a}_{\mathbf{j}}, \xi)) = \mathbf{a}_{\mathbf{j}}\}} \right. \\ & \left. \times \prod_{j \in \mathbf{j} + \mathbf{t} - 1} n^{-1} S_j \prod_{j \in \{0, \dots, n\} \setminus \mathbf{j} + \mathbf{t} - 1} p_j(\xi) \right) \quad (\text{B.1}) \end{aligned}$$

where the Jacobian

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

*Proof.* If  $F(F^{-1}(\xi, u), \xi) = u$  for each fixed  $\xi$ , then the lemma follows immediately from 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  $\tilde{F}(a_i, \xi) = F(a_i, \xi)$  for all  $i = 1, \dots, k$  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  $\mathbf{a}, \mathbf{b} \in \{a_0, \dots, a_{k+1}\}$ , 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 fiducial distributions (2.2) computed based on the structural equation for  $\mathbf{X}$  and  $\tilde{\mathbf{X}}$  are the same.  $\square$

*Proof Theorem 5.1.* We will prove the theorem in two steps. First, we prove Bernstein-von Mises for some special points in the  $Q((\mathbf{a}, \mathbf{b}), u)$  and verify the conditions of Theorem 1 of Hannig (2009) for them. We only need to verify Assumptions 1 and 2 as the 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  $Q((\mathbf{a}, \mathbf{b}), u)$ .

Define  $\mathbf{p} = (p_0(\xi_0), \dots, p_k(\xi_0))$  and  $\Sigma$  the variance matrix of the Multinomial( $1, \mathbf{p}$ ) distribution. By the Skorokhod representation theorem we can find  $\mathbf{S}$  having Multinomial( $n, \mathbf{p}$ ) distribution and  $\mathbf{H}$  having Normal( $0, \Sigma$ ) distribution such that  $\mathbf{S} = n\mathbf{p} + n^{\frac{1}{2}}\mathbf{H} + o_{as}(n^{\frac{1}{2}})$ ,  $n \rightarrow \infty$ . 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  $n^{\frac{1}{2}}(\mathcal{R}_\xi^{\mathbf{d}} - \xi_0)$  is  $g(\mathbf{z}) = r_{\mathbf{d}}(\xi_0 + n^{-\frac{1}{2}}\mathbf{z})n^{-\frac{p}{2}}$ . We will investigate the behavior of  $g(\mathbf{z})$  as  $n \rightarrow \infty$ .

First set

$$g_{2,\mathbf{j}}(\mathbf{z}) = J(\mathbf{a}_{\mathbf{j}}, \xi) \left( \sum_{\mathbf{t} \in \{0,1\}^p} I_{\{\mathbf{c}_{\mathbf{j},\mathbf{t}}^{\mathbf{d}}(\mathbf{G}_{\mathbf{j}}^{-1}((\mathbf{a}_{\mathbf{j}+\mathbf{t}-1}, \mathbf{a}_{\mathbf{j}+\mathbf{t}}), \xi)) = \mathbf{a}_{\mathbf{j}}\}} \right) \\ \times \prod_{j \in \mathbf{j} + \mathbf{t} - 1} n^{-1} S_j \prod_{j \in \{0, \dots, n\} \setminus \mathbf{j} + \mathbf{t} - 1} p_j(\xi).$$

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

Second, compute using Taylor series and Stirling's formula

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

By our assumptions

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

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

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

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

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

By plugging back into (B.2) we get

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

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

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

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

the diagonal  $k \times k$ -matrix  $D = \text{diag}(p_1(\xi_0), \dots, p_k(\xi_0))^{-1}$  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. Moreover, 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  $V = A \tilde{\Sigma}^{-1} A^\top$ . Also 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  $n(\mathbf{z}) = C \tilde{n}(\mathbf{z})$ , with the constant

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

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

$$g_1(\mathbf{z}) J((a_1, \dots, a_p), \xi_0 + n^{-\frac{1}{2}} \mathbf{z}) \rightarrow n(\mathbf{z}) \quad a.s. \quad (\text{B.4})$$

However, since the function  $g_1(\mathbf{z}) J((a_1, \dots, a_p), \xi_0 + n^{-\frac{1}{2}} \mathbf{z})$  is a transformation of Dirichlet density, it integrates to 1. Hence the convergence in (B.4) is also in  $L^1$ . Since  $0 \leq n^{-1} S_j \leq 1$  and  $0 \leq p_j(\xi) \leq 1$ , the uniform integrability of  $g_1(\mathbf{z}) g_{2,\mathbf{j}}(\mathbf{z})$  follows and one can conclude that  $n^{\frac{1}{2}} (\mathcal{R}_\xi^{\mathbf{d}} - \xi_0)$  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  $\mathbf{j}$ ,  $g_1(\mathbf{z}) g_{2,\mathbf{j}}(\mathbf{z}) \rightarrow D_{\mathbf{j}} n(\mathbf{z})$  *a.s.*, and

$$g_1(\mathbf{z}) g_{2,\mathbf{j}}(\mathbf{z}) \leq 2^k g_1(\mathbf{z}) J(\mathbf{a}_{\mathbf{j}}, \xi_0 + n^{-\frac{1}{2}} \mathbf{z}).$$

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

$$\tilde{p}_{\mathbf{j}}(\xi) = p_{i_{\mathbf{j}}}(\xi) + \dots + p_{i_{\mathbf{j}+1}-1}(\xi), \quad \tilde{S}_{\mathbf{j}} = S_{i_{\mathbf{j}}} + \dots + S_{i_{\mathbf{j}+1}-1},$$



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

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

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

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

Thus  $g_1(\mathbf{z})g_{2,j}(\mathbf{z})$  is uniformly integrable and  $g_1(\mathbf{z})g_{2,j}(\mathbf{z}) \rightarrow D_j n(\mathbf{z})$  in  $L^1$ . We conclude by a straightforward calculation  $n^{\frac{1}{2}}(\mathcal{R}_\xi^d - \xi_0)$  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  $\text{Var}(V^{-1}A\tilde{\Sigma}^{-1}\tilde{\mathbf{H}}) = V^{-1}$ . The Assumptions 1, 2 of Theorem 1 in Hannig (2009) are verified for the special extreme points  $\mathcal{R}_\xi^d$ . The first step of the proof is complete.

Now we will finish the proof by showing that

$$\text{diam } Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) \mid \{Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) \neq \emptyset\} = O_P(n^{-1}) \quad a.s. \quad (\text{B.5})$$

This, together with what was proved above, will verify the Assumptions 1, 2 of Theorem 1 in Hannig (2009) for  $\mathcal{R}_\xi$  based on any point in  $Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*)$ .

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

$$\begin{aligned} P(\text{diam } Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) > K/n \mid Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) \neq \emptyset) \\ = \int P(\text{diam } Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) > K/n \mid Q_d((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) = \xi) f_{\mathcal{R}^d}(\xi), \end{aligned}$$

where  $f_{\mathcal{R}^d}(\xi)$  is the density of  $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{U}^*)$  given  $\{Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) \neq \emptyset\}$  displayed in (B.1). For  $\mathbf{i} \subset \{1, \dots, n\}$  denote by  $J_{\mathbf{i}}^d$  the event that the  $Q_d((\mathbf{a}, \mathbf{b}), \mathbf{U}^*)$  is based of the  $p$  observational inequalities for  $\mathbf{X}_{\mathbf{i}}$ . We then have

$$\begin{aligned} P(\text{diam } Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) > K/n \mid Q_d((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) = \xi) \\ = \sum_{\mathbf{i}} P(\text{diam } Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) > K/n \mid J_{\mathbf{i}}^d, \{Q_d((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) = \xi\}) \\ \times P(J_{\mathbf{i}}^d \mid Q_d((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) = \xi) \end{aligned}$$

Consider  $J_{\mathbf{i}}^d$ . The observational inequalities labeled by  $a_{k_i} < x_i \leq a_{k_i+1}$ ,  $i \in \mathbf{i}$  are used for computing  $Q_d$ . From the remaining observational inequalities we get  $\mathbf{U}_{\mathbf{i}^c}^* \mid J_{\mathbf{i}}^d \cap \{Q_d((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) = \xi\}$  are independent Uniform( $F(a_{k_j}, \xi)$ ,  $F(a_{k_j+1}, \xi)$ ) random variables, respectively, i.e., conditionally on  $J_{\mathbf{i}}^d \cap \{Q_d((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) = \xi\}$ , the random variables  $\tilde{U}_i^* = (U_i^* - F(a_{k_i}, \xi)) / (F(a_{k_i+1}, \xi) - F(a_{k_i}, \xi))$ ,  $i \in \mathbf{i}^c$ , are

i.i.d. Uniform(0, 1). For each  $j = 0, \dots, k$ , 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  $J_{\mathbf{i}}^{\mathbf{d}} \cap \{Q_{\mathbf{d}}((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) = \xi\}$  we have

$$Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) \subset \left( \bigcap_{j=1}^k \{\bar{\xi} : F(a_j, \bar{\xi}) \leq F(a_{k_j}, \xi) + (F(a_{k_{j+1}}, \xi) - F(a_{k_j}, \xi)) \tilde{U}_{1:S_{j-1}}^j\} \right. \\ \left. \cap \bigcap_{j=0}^{k-1} \{\bar{\xi} : F(a_{j+1}, \bar{\xi}) \geq F(a_{k_{j+1}}, \xi) + (F(a_{k_{j+1}}, \xi) - F(a_{k_j}, \xi))(1 - \tilde{U}_{S_{j-1}:S_{j-1}}^j)\} \right) \quad (\text{B.6})$$

Here  $\tilde{U}_{1:S_{j-1}}^j$  and  $\tilde{U}_{S_{j-1}:S_{j-1}}^j$  are the order statistics of an array obtained by reordering  $\tilde{U}_{\mathbf{i}}^*$ ,  $\mathbf{i} \in \mathbf{i}^{\mathbf{c}}$ , 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  $\xi_0$ , there is an open neighborhood  $\mathcal{N}$  of  $\xi_0$  and a constant  $C$  such that for all  $\xi \in \mathcal{N}$  and all  $\mathbf{i}$ ,

$$\text{diam } Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) \leq C \max\{\tilde{U}_{1:S_{j-1}}^j, 1 - \tilde{U}_{S_{j-1}:S_{j-1}}^j, j = 0, \dots, k\}.$$

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

$$P(\text{diam } Q((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) > K/n \mid J_{\mathbf{i}}^{\mathbf{d}}, Q_{\mathbf{d}}((\mathbf{a}, \mathbf{b}), \mathbf{U}^*) = \xi) \geq \epsilon$$

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

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

## References

- Barnard, G. A. (1995) Pivotal Models and the Fiducial Argument. *International Statistical Reviews*, **63**, 309–323.
- Casella, G. and Berger, R. L. (2002) *Statistical Inference*. Pacific Grove, CA: Wadsworth and Brooks/Cole Advanced Books and Software, 2nd edn.
- Dawid, A. P. and Stone, M. (1982) The functional-model basis of fiducial inference. *Ann. Statist.*, **10**, 1054–1074. With discussions by G. A. Barnard and by D. A. S. Fraser, and a reply by the authors.
- Dempster, A. P. (2008) The Dempster-Shafer Calculus for Statisticians. *International Journal of Approximate Reasoning*, **48**, 365–377.
- Denoeux, T. (2006) Constructing belief functions from sample data using multinomial confidence regions. *International Journal of Approximate Reasoning*, **42**, 228–252.

- E, L., Hannig, J. and Iyer, H. K. (2008) Fiducial Intervals for Variance Components in an Unbalanced Two-Component Normal Mixed Linear Model. *Journal of American Statistical Association*, **103**, 854–865.
- (2009) Fiducial Generalized Confidence interval for Median Lethal Dose (LD50). Submitted.
- Fisher, R. A. (1930) Inverse probability. *Proceedings of the Cambridge Philosophical Society*, **xxvi**, 528–535.
- Fraser, D. A. S. (1968) *The Structure of Inference*. New York-London-Sydney: John Wiley & Sons Inc.
- Hannig, J. (2009) On Generalized Fiducial Inference. *Statistica Sinica*, **19**, 491–544.
- Hannig, J., Iyer, H. K. and Patterson, P. (2006) Fiducial generalized confidence intervals. *Journal of American Statistical Association*, **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.
- (2009) Prediction in Fiducial Inference. Preprint.
- Hannig, J. and Lee, T. C. M. (2009) Generalized fiducial inference for wavelet regression. *Biometrika*. To appear.
- Lindley, D. V. (1958) Fiducial distributions and Bayes' theorem. *J. Roy. Statist. Soc. Ser. B*, **20**, 102–107.
- 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. (2009) Fiducial Inference on the Maximum Mean of a Multivariate Normal Distribution. Preprint.
- Wang, J. C.-M. and Iyer, H. K. (2006) A generalized confidence interval for a measurand in the presence of type-A and type-B uncertainties. *Measurement*, **39**, 856–863.
- Weerahandi, S. (1993) Generalized confidence intervals. *J. Amer. Statist. Assoc.*, **88**, 899–905.
- Zabell, S. L. (1992) R. A. Fisher and the fiducial argument. *Statist. Sci.*, **7**, 369–387.