

INTRODUCTION TO GENERALIZED FIDUCIAL INFERENCE

BY ALEXANDER MURPH¹, JAN HANNIG¹ AND JONATHAN P WILLIAMS²

¹*Department of Statistics & Operations Research, University of NC at Chapel Hill, acmurph@live.unc.edu; jan.hannig@unc.edu*

²*Department of Statistics, NC State University, Raleigh, NC, jwilli27@ncsu.edu*

Fiducial inference was introduced in the first half of the 20th century by Fisher (1935) as a means to get a posterior-like distribution for a parameter without having to arbitrarily define a prior. While the method originally fell out of favor due to non-exactness issues in multivariate cases, the method has garnered renewed interest in the last decade. This is partly due to the development of generalized fiducial inference, which is a fiducial perspective on generalized confidence intervals: a method used to find approximate confidence distributions. In this chapter, we illuminate the usefulness of the fiducial philosophy, introduce the definition of a generalized fiducial distribution, and apply it to interesting, non-trivial inferential examples.

1. Introduction. Fiducial inference attempts to find a middle ground between frequentist and Bayesian perspectives. The fiducial argument allows one to fit a posterior-like distribution on a target parameter θ in a way that is entirely data-driven and does not rely on a sometimes arbitrary prior selection. This argument is based on inverting a *data generating algorithm* (DGA) that associates data \mathbf{Y} to the parameters θ and a random component U with a known distribution F_0 , e.g. a vector of i.i.d. of standard uniforms or standard Gaussians. This is often expressed as the relation $\mathbf{Y} = A(U, \theta)$. Since the DGA is a function of a known distribution, it immediately determines the likelihood $f(\mathbf{y}|\theta)$. By solving the DGA for θ , we get the distribution of our parameters, called the *generalized fiducial distribution* (GFD), that is entirely data-driven and does not require the use of Bayes' Theorem.

While the strengths and limitations of *generalized fiducial inference* (GFI) continue to be explored, its usefulness has already been illustrated in numerous practical applications. Recent work has applied the fiducial ideas to bio-equivalence (McNally *et al.*, 2003; Hannig *et al.*, 2003), metrology problems (Hannig *et al.*, 2013; Wang and Iyer, 2006a,b; Hannig *et al.*, 2007; Wang *et al.*, 2012), inter-laboratory experiments and international key comparison experiments (Iyer *et al.*, 2004). It has also been used to tackle statistical problems at the forefront of modern topics in statistical research, such as in wavelet regression (Hannig and Lee, 2009) and extreme value estimation (Wandler and Hannig, 2012a), and has recently led to a creative new perspective on linear model selection (Williams and Hannig, 2019) and vector autoregressive graph selection (Williams *et al.*, 2019).

Formally, GFD is defined as a limit (Hannig *et al.*, 2016) in the following way. Start by defining the inverse of the DGA using the optimization problem

$$(1) \quad Q_{\mathbf{y}}(u) = \operatorname{argmin}_{\theta^*} \|\mathbf{y} - A(u, \theta^*)\|.$$

Typically $\|\cdot\|$ is either ℓ_2 or ℓ_∞ norm. Next, for each small $\epsilon > 0$, define the random variable $\theta_\epsilon^* = Q_{\mathbf{y}}(U_\epsilon^*)$, where U_ϵ^* has distribution F_0 **truncated** to the set

$$(2) \quad \mathcal{M}_{\mathbf{y}, \epsilon} = \{U_\epsilon^* : \|\mathbf{y} - A(U_\epsilon^*, \theta_\epsilon^*)\| = \|\mathbf{y} - A(U_\epsilon^*, Q_{\mathbf{y}}(U_\epsilon^*))\| \leq \epsilon\}.$$

Then assuming that the random variable θ_ϵ^* converges in distribution as $\epsilon \rightarrow 0$, the GFD is defined as the limiting distribution of $\lim_{\epsilon \rightarrow 0} \theta_\epsilon^*$.

When the sampling distribution of \mathbf{Y} is discrete, we can set $\epsilon = 0$ and no limit is necessary. When the sampling distribution of \mathbf{Y} is continuous Hannig *et al.* (2016) show the following result.

THEOREM 1.1. *Under mild conditions, the limiting distribution above has density*

$$(3) \quad r_{\mathbf{y}}(\theta) = \frac{f(\mathbf{y}|\theta)J(\mathbf{y},\theta)}{\int f(\mathbf{y}|\theta')J(\mathbf{y},\theta')d\theta'}$$

where $J(\mathbf{y},\theta) = D\left(\nabla_{\theta}A(u,\theta)|_{u=A^{-1}(\mathbf{y},\theta)}\right)$. Here $\nabla_{\theta}A(u,\theta)$ is a gradient matrix computed with respect to θ , and D is a determinant like operator that depends on the norm in (1), e.g., when we use ℓ_2 norm $D(M) = (\det M'M)^{\frac{1}{2}}$.

The generalized fiducial approach communicates a simple algorithm: when possible, define a DGA that expresses the relationship between your data, your parameter, and a random quantity, then invert it. The application of this idea in practice, however, can be nuanced and requires careful thought. Our aim in this chapter is to provide a comprehensive overview of GFI by way of numerous detailed examples. A common theme among our examples, and indeed a unifying aspect of all fiducial solutions, is the development of inferentially meaningful subsets of the parameter space. As we will illustrate below, inverting a DGA does not lead to a single point estimate, but instead a set of possible values, which then becomes a distribution of sets, inherited from the random quantity U .

In particular, we will treat two instances of multivariate normal data and a binomial distribution with unknown number of trials. We selected these examples both because they are of interest in their own right, but also because they will allow us to demonstrate how to implement a generalized fiducial solution using modern computational tools. The computer codes are available at a GitHub repository [<https://github.com/sirmurphalot/IntroductionGFI>].

2. Multivariate Normal Distribution.

2.1. *Multivariate Normal Data; μ unknown, Σ unknown.* The estimation of covariance matrices is a fundamental problem in many multivariate methods. Examples include discriminant data analysis, longitudinal data analysis, time series analysis and spatial data analysis, just to name a few. However, only recently it was pointed out that in the Bayesian context the most commonly used conjugate inverse Wishart prior has a problem (Berger *et al.*, 2020a,b; Yang and Berger, 1994). In particular, the inverse Wishart posterior has the effect of forcing the eigenvalues of the covariance matrix apart. This is quite undesirable, because in practice it is reasonable to expect that the true covariance matrix has some or even all eigenvalues equal to each other.

Let $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m$ be iid $\mathcal{N}_d(\mu, \Sigma)$, where μ is a d -dimensional vector and Σ is a $d \times d$ covariance matrix. Our aim is to perform inference on the covariance matrix Σ with the unknown mean parameter μ . To date there have been proposed two basic approaches to define the GFD for this model. They both start with the DGA,

$$\mathbf{Y}_i = \mu + B\mathbf{Z}_i, \quad i = 1, \dots, n,$$

where \mathbf{Z}_i are i.i.d. standard Gaussian vectors of dimension d , but differ in the structure of the matrix B . Wandler and Hannig (2011) use B that is a lower triangular matrix. This leads to GFD that depends on the arbitrary order of the coordinates. Shi *et al.* (2017) propose using an arbitrary $d \times d$ matrix B , which removes the dependence on the coordinate order but is overparametrized. The resulting GFD for the covariance matrix $\Sigma = BB^{\top}$ belongs to the Wishart family.

To address these issues, we propose an alternative DGA that does not lead to the Wishart distribution. In particular, consider $B = U\Lambda$, where U is an orthogonal matrix, Λ is a diagonal matrix with positive entries on the diagonal. Consequently, the covariance matrix is $\Sigma =$

$U\Lambda^2U^\top$. To be able to compute the GFD using (3) we will need to reparametrize U using the *Cayley transformation* (see Theorem 2.1). In particular we will use the following two facts. For the proof of the first fact see, for instance, [Eves \(1996\)](#).

THEOREM 2.1 (Cayley transform). *Every real orthogonal matrix U that does not have -1 as a characteristic root can be expressed as*

$$(4) \quad U = (I_d + A)(I_d - A)^{-1} = (I_d - A)^{-1}(I_d + A)$$

by a suitable choice of a real skew-symmetric matrix, i.e., $A^\top = -A$.

THEOREM 2.2 ([O'Dorney \(2014\)](#)). *For any orthogonal matrix U there must exist a signature matrix D such that UD does not have -1 as a characteristic root and all elements the of corresponding Caley transform $a_{i,j}$ are such that $|a_{i,j}| \leq 1$ for $1 \leq i \leq j \leq d$.*

Recall that D is a diagonal matrix with entries ± 1 . Consequently, $UDA^2DU^\top = U\Lambda^2U^\top$ which is invariant to the choice of D . This leads us to propose the following data generating algorithm

$$(5) \quad \mathbf{Y}_i = \mu + (I_d + A)(I_d - A)^{-1}\Lambda \mathbf{Z}_i, \quad i = 1, \dots, n,$$

where A is a skew-symmetric matrix with all entries $|a_{ij}| \leq 1$ and Λ is a diagonal matrix with positive entries $\lambda_i > 0$.

We show in Appendix A that the Jacobian is $J(\mathbf{y}, \theta) = J^*(\mathbf{y}, A) \prod_{i=1}^n \lambda_i^{-1}$, where $J^*(\mathbf{y}, A)$ does not depend on μ and λ . The form of this Jacobian allows us to simplify the generalized fiducial density from (3). In particular, calculations in Appendix A show that the marginal GFD of A is

$$(6) \quad r_{\mathbf{y}}(A) \propto J^*(\mathbf{y}, A) \prod_{i=1}^d (U^T SSE U)_{ii}^{\frac{-(n-1)}{2}},$$

where $SSE = \sum_{i=1}^n (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^\top$, $\bar{\mathbf{Y}} = n^{-1} \sum_{i=1}^n \mathbf{Y}_i$, and U is defined in (4). The conditional GFD of the entries of Λ^{-2} given A is independent of $\text{Gamma}\left(\frac{n-1}{2}, \frac{(U^T SSE U)_{ii}}{2}\right)$, $i = 1, \dots, d$. The conditional GFD of μ given Λ and A is multivariate $N(\bar{\mathbf{Y}}, n^{-1} U \Lambda^2 U^\top)$. The simple form of this distribution allows us to implement sampling from the GFD using STAN ([Stan Development Team, 2020](#)).

Since A is skew-symmetric, its diagonal terms must be zero and there are only $d(d-1)/2$ unique elements, which we take to correspond to the number of elements in the lower triangle of A . The number of free parameters (A, Λ, μ) in (5) is therefore $d(d+3)/2$, which matches the number of free equations in the minimal sufficient statistic $(\bar{\mathbf{Y}}, SSE)$. Consequently, our DGA is not overparameterized. In order to be able to efficiently marginalize the GFD, we do not impose an order restriction on the diagonal entries of Λ that introduces non-uniqueness due to the order of λ_i . To address this issue, we run parallel chains of our MCMC algorithm each starting with their own random ordering of the singular values, obtained by using PCA on the sample covariance matrix.

The STAN software ([Stan Development Team, 2020](#)) allows us to quickly and easily draw samples from the GFD, which we can then use to obtain approximate confidence intervals for the true covariance matrix. To build these confidence intervals and evaluate our method, we consider a number of distance metrics and parameters of interest. For the distance metrics, confidence sets are developed by defining a ball around the mean of the GFD that covers $(1 - \alpha)\%$ of the GFD. For parameters of interest, confidence sets are developed much like they

would be on the real line. That is, after mapping every matrix from the GFD to the real line, we simply take the center $(1 - \alpha)\%$ of the GFD as our confidence set.

In particular, we construct the confidence sets for the true covariance matrix using the following two distance metrics, two norms, and one parameter of interest. The norms are used as both distance metrics (for matrices $M, N : ||M - N||$) and as parameters of interest (for matrix $M : ||M||$). Let M, N be two positive-definite, symmetric matrices. Then,

1. FM_DISTANCE (Förstner and Moonen, 2003) : $dist(M, N) = \sqrt{\sum_{i=1}^d \ln^2 \lambda_i(M, N)}$, where $\lambda_i(M, N)$ are the eigenvalues from the equation $\det(\lambda M - N) = 0$;
2. STEIN_LOSS (Konno, 1995) : $dist(M, N) = \text{tr}(M^{-1}N) - \log(\det(M^{-1}N)) - d$;
3. SPECTRAL_NORM (Horn and Johnson, 2012) : $norm(M) = \lambda_{(d)}$, the maximum value in the Λ matrix;
4. FROBENIUS_NORM (Horn and Johnson, 2012) : $norm(M) = \sqrt{\text{tr}(MM^H)}$, where M^H is the conjugate transpose of H ;
5. LOGDET (Fazel *et al.*, 2003) : $huer(M) = \log(\det(M))$.

Our simulation study consists of 1000 iterations on a 4×4 multivariate normal $\mathcal{N}_4(\mu, \Sigma)$ where

$$\mu = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 4 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 9 & 1 \\ 0 & 1 & 1 & 4 \end{pmatrix}.$$

At each iteration, we simulate a dataset of $n = 100$ observations from the given multivariate normal distribution. Then, we run 20 MCMC chains and use the simulated values from the GFD of the covariance matrix to construct approximate confidence intervals around the true covariance matrix Σ .

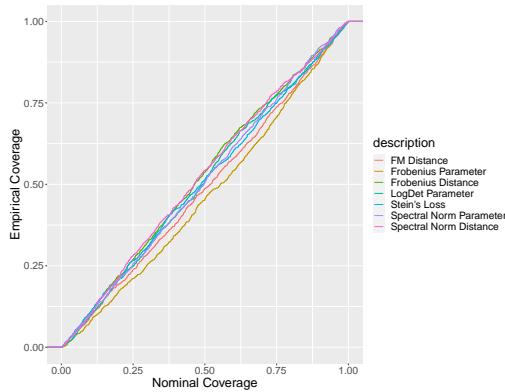


FIG 1. The coverage of confidence sets for the covariance matrix using different notions of distance. This figure is created by setting different levels of coverage in our GFD credible sets and checking the empirical coverage of these intervals over the whole simulation. A perfect diagonal line equates to a match between our observed (empirical) coverage and our theoretical (nominal) coverage.

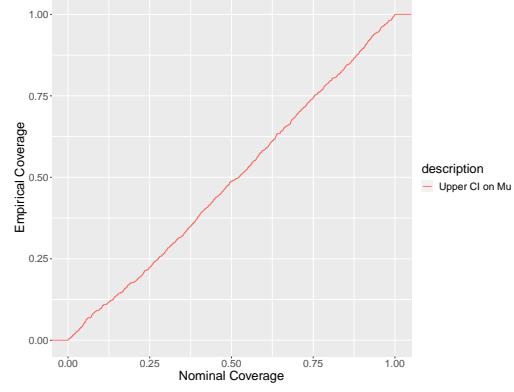


FIG 2. The coverage of credible intervals for the mean vector μ using standard euclidean distance. This figure was made in the same was as in figure 1.

Figure 1 shows the performance of these confidence sets for different values of α . The nominal coverage corresponds to the value of α we set and the empirical coverage is the

Distance Metric/Norm	FM_DISTANCE
FM_DISTANCE	0.9505
LOSS1_DISTANCE	0.9463
LOGDET Parameter	0.9589
SPECTRAL_NORM Parameter	0.9568
FROBENIUS_NORM Parameter	0.9547
SPECTRAL_NORM Distance	0.9540
FROBENIUS_NORM Distance	0.9580

TABLE 1
The coverage of credible intervals for the covariance matrix using different notions of difference.

proportion of times $(1 - \alpha)$ balls made this way contained the true covariance matrix. As we can see, all of our metrics resulted in a line across the diagonal, communicating that nominal and empirical coverage match. Table 1 shows the specific coverage of these confidence sets for $\alpha = 0.05$. As we can see, the empirical coverage probabilities of each of these intervals is approximately 95%, as desired.

The simultaneous confidence set for μ is constructed the same way as Σ using the standard Euclidean distance $dist(u, v) = ||u - v||$. Draws from the GFD for μ were calculated at the same time as the draws of Σ in our simulation study. Figure 2 shows that nominal coverage matched our empirical coverage, validating our method in the context of this inference problem.

2.2. Generalization of one way random effects model. Continuing to examine the multivariate normal problem, we consider an alternative, more restrictive parameterization, standard unbalanced one-way random model $Y_{i,j} = \mu + \eta_i + \epsilon_{i,j}$, for $\eta_i \sim \mathcal{N}(0, \sigma_\alpha^2)$, $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma_e^2)$. The first fiducial solution to inference on this model was given by [E et al. \(2008\)](#) and used a tailor made solution to this problem. Let us consider the more general DGA

$$(7) \quad \mathbf{Y} = \mathbb{X}\beta + A\mathbf{Z},$$

where β is an unknown matrix of fixed effects, \mathbf{Z} is standard normal random vector, \mathbb{X} is a fixed effect design matrix and $A = \Sigma^{1/2}$ such that $\Sigma = \sigma_\alpha^2 S_\alpha + \sigma_e^2 I$ and S_α is a matrix of ones and zeros corresponding to the group sizes n_1, n_2, \dots, n_m . Thus, our inferential problem is simplified to providing GFD for β , σ_α^2 , and σ_e^2 .

While [E et al. \(2008\)](#) were able to derive well-performing generalized fiducial intervals for σ_α^2 and σ_e^2 , they did so by way of a long calculation and they did not allow for simultaneous inference on the fixed effects β . Using (3) the generalized fiducial intervals can be implemented without major computational hassle using the STAN software ([Stan Development Team, 2020](#)), while simultaneously obtaining GFD for the fixed effects β that is often of interest ([Neupert et al., 2020](#)).

Due to the structure imposed by (7), the Jacobian matrix simplifies greatly and does not involve the square root of a complicated matrix. Indeed, the function $J(\mathbf{Y}, S_\alpha, \sigma_\alpha^2, \sigma_e^2, \mathbb{X}\beta)$ is computed from the $n \times (d + 2)$ matrix obtained by column concatenation of

$$\nabla_\beta \mathbf{Y} = \mathbb{X}, \quad \frac{\partial \mathbf{Y}}{\partial \sigma_e^2} = (\sigma_\alpha^2 S_\alpha + \sigma_e^2 I)(\mathbf{Y} - \mathbb{X}\beta), \quad \frac{\partial \mathbf{Y}}{\partial \sigma_\alpha^2} = S_\alpha (\sigma_\alpha^2 S_\alpha + \sigma_e^2 I)(\mathbf{Y} - \mathbb{X}\beta).$$

Derivation of the above quantities is outlined in Appendix B.

We will draw samples from our GFD to create credible sets that we will then compare to the truth to assess performance, parallel the simulation study in [E et al. \(2008\)](#). In particular, at each instance of our simulation, we generated data using seven different designs of group sizes outlined in Table 2 and one of the true parameter values $(\sigma_\alpha^2, \sigma_e^2)$ from the set $\{(1, 10), (.5, 10), (1, 10), (.5, 2), (1, 1), (2, .5), (5, .2), (10, .1)\}$.

Pattern	Φ	n_i					
1	.068	1	1	1	1	1	100
2	.130	2	2	2	2	2	100
3	.187	2	5	60			
4	.410	4	4	4	8	48	
5	.700	5	10	15	20	25	30
6	.807	2	2	4	6		
7	.957	6	6	8	8	10	10

TABLE 2

The different group assignments tested for simulation of mixed models, as suggested by [E et al. \(2008\)](#). Here, Φ is an expression of imbalance as defined by [Ahrens and Pincus \(1981\)](#). The smaller the value of Φ , the greater the degree of imbalance.

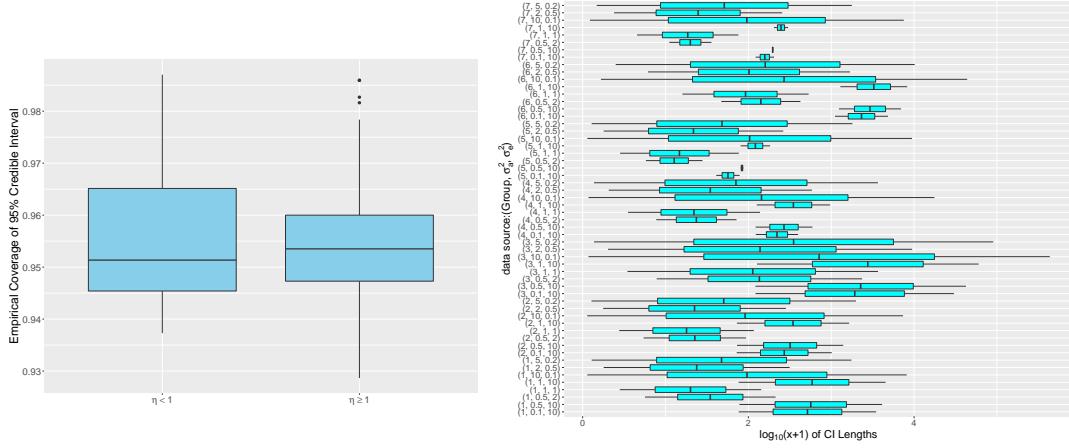


FIG 3. Empirical coverages for $\eta = \sigma_\alpha^2/\sigma_e < 1$ and $\eta \geq 1$ for both σ_α^2 and σ_e^2 across all parameter and group choices. We can see that the generalized fiducial method tends to be more conservative for larger values of η .

FIG 4. $\log_{10}(x+1)$ of the CI lengths for all our parameters. Instances of outstanding lengths likely contribute to the conservative empirical coverage we see in 3

The results of our simulation can be seen in Figures 3 and 4. Across many different parameter choices, it would appear that the generalized fiducial method performs well, with a slight conservative tendency for certain extreme parameter combinations. Our findings match [E et al. \(2008\)](#), who found this same conservative tendency for larger values of $\eta = \sigma_\alpha^2/\sigma_e^2$ and greater degrees of imbalance, yet promising performance in all other instances. We conclude by pointing out that we were able to efficiently implement our calculations using RStan ([Stan Development Team, 2020](#)) which yields accurate results with a parsimonious amount of iterations.

3. Binomial Distribution.

3.1. *Binomial Distribution; n known, p unknown.* When working with discrete distributions, Equation (3) can no longer be used, and one must instead invert the DGA directly. To demonstrate how one applies the generalized fiducial method to discrete data, we begin with a well studied application of the GFI ideas: inference on p from the binomial distribution with n known. The following is a possible data generating algorithm

$$(8) \quad Y = \sum_{i=1}^n I_{(0,p)}(U_i),$$

where $p \in [0, 1]$ is the unknown parameter and U_i are i.i.d. Uniform(0, 1).

Assume we have an observed realization of our random variable $Y = y$. The uniform random variables U_i that generated the observation y are unknown and therefore we replace them with newly generated U_i^* . When we solve (8) for p , we obtain the interval $(U_{(y)}^*, U_{(y+1)}^*]$, where $U_{(y)}^*$ is the y th order statistic of U_i^* . Recall that $U_{(y)}^*$ follows the $Beta(y, n - y + 1)$ distribution

This illustrates an important feature of GFI for discrete data: it is based on a random set rather than a random point. Notions from Dempster-Shafer theory (Dempster, 2008; Edlefsen et al., 2009), such as *Belief* and *Plausibility*, can be used to interpret these sets. Another approach is to reduce the interval to a single distribution. Typically this is defined as an average of the densities of the upper and lower bound of the interval, the density of $Beta(y, n - y + 1)$ and $Beta(y + 1, n - y)$. Hannig (2009) suggest using the arithmetic mean, while Schweder and Hjort (2016) advocate for the geometric mean. A 1/3 mean of the beta densities has been suggested based on higher order asymptotics in an unpublished manuscript. A simple calculation yields that

$$\begin{aligned} r_{y,n}(p) &\propto (\sqrt[3]{Beta(y+1, n-y)} + \sqrt[3]{Beta(y, n-y+1)})^3 \\ &\propto Beta(y+1, n-y) + c_1 Beta(y + \frac{2}{3}, n-y + \frac{1}{3}) \\ &\quad + c_2 Beta(y + \frac{1}{3}, n-y + \frac{2}{3}) + Beta(y, n-y+1), \\ c_1 &= 3 \left(\frac{\Gamma(y+2/3)\Gamma(n-y+1/3)}{y^{2/3}(n-y)^{1/3}\Gamma(y)\Gamma(n-y)} \right), \quad c_2 = 3 \left(\frac{\Gamma(y+1/3)\Gamma(n-y+2/3)}{y^{1/3}(n-y)^{2/3}\Gamma(y)\Gamma(n-y)} \right). \end{aligned}$$

Thus, simulating values from this GFD is computationally easy and effective, as it is just a mixture of Beta distributions.

This result immediately generalizes to multiple independent observations from a binomial distribution. Let Y_1, \dots, Y_m be i.i.d. $Bin(n, p)$ where n is known but p is unknown. The sufficient statistic $\sum_{i=1}^m Y_i$ follows $Bin(nm, p)$ and we can use (8) to obtain a GFD.

3.2. Binomial Distribution; n unknown, p known. With the aim of providing a more nuanced perspective on the generalized fiducial approach for discrete data, we will now consider a more complicated problem: binomial data for n unknown. To the best of our knowledge, the binomial distribution with unknown n has yet to receive a fiducial solution in the literature.

Consider m i.i.d. observations from a $Bin(n, p)$ distribution, $\mathbf{Y} = (Y_1, \dots, Y_m)$ where p is known but n is unknown, and our objective is inference on n . We begin our solution by choosing a DGA based on the inverse of the distribution function. In particular,

$$(9) \quad Y_i = F_{n,p}^{-1}(U_i), \quad i = 1, \dots, m,$$

where U_1, \dots, U_m are i.i.d. $Unif(0, 1)$, and $F_{n,p}^{-1}(u) = \inf\{y : F_{n,p}(y) \geq u\}$ with $F_{n,p}$ being the distribution function of $Bin(n, p)$.

There are two major observations to make about inverting (9): first, that $n \geq \max\{\mathbf{Y}\}$; and second, that n does not have an upper bound. Because of this latter observation, we must begin by approximating the GFD by restricting our calculations to a range of reasonable values $\{\max\{\mathbf{Y}\}, \dots, n-1, n\}$. Then, the generalized fiducial probability mass of all relevant subsets in this range is determined by a deterministic algorithm. That is, unlike the usual simulation based algorithms this algorithm directly assigns influentially meaningful probabilities to sets of possible n values. The precise details of this algorithm can be found in Appendix C.

We remark that the difference between GFD and Bayesian posterior using the flat prior is a result of the fiducial probability assigned to non-singleton sets, and if that probability is negligible the GFD would coincide with the Bayesian posterior. In the simulations below, the final GFD on the values of n is obtained by sampling from end-points of all of these sets that have positive fiducial probability.

We illustrate the generalized fiducial solution to this inference problem via a simulation study, comparing the generalized fiducial solution against the Bayesian. For this simulation, we generate data with 10, 50 and 100 observations from the distribution $\text{Bin}(n_0, p)$ where $n_0 = 10$ always and p varies over the set $\{0.01, 0.05, 0.1, 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99\}$. For each instance of our binomial data we draw 1000 values \hat{n}_B from the Bayesian posterior distribution on n as well as 1000 values \hat{n}_F from the GFD on n . Figure 5 shows the 95% credible intervals based off of these 1000 draws. This simulation reveals that while the generalized fiducial solution does capture the true value n_0 with an expected amount of regularity, there is not much distinguishing it from the Bayesian solution to the same problem.

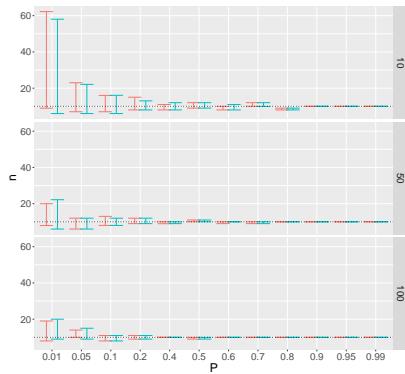


FIG 5. 95% credible intervals for each candidate value of p .

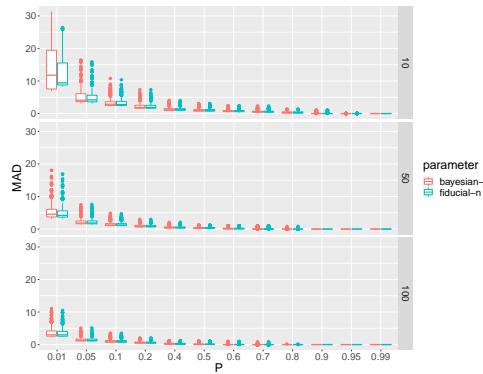


FIG 6. Boxplots of 1000 MAD calculations over varying values of p and data sizes.

To better examine the difference between the generalized fiducial and the Bayesian solutions to this problem, we perform the above simulation 1000 times: each time drawing 1000 values from the GFD and 1000 values from the Bayesian posterior distribution. At each repetition, we record the Mean Absolute Difference (MAD) $\frac{1}{1000} \sum_{i=1}^{1000} |n_0 - \hat{n}_{\eta,i}|$ for both the generalized fiducial and the Bayesian draws, where $\eta \in \{\text{Fiducial, Bayesian}\}$.

Figure 6 compares boxplots of each method's 1000 MAD values collected via the described simulation. We can see from this simulation that, while the generalized fiducial and Bayesian solutions become more alike for larger data sizes, they are different for smaller data sizes and for smaller values of p . In particular, for the draw of 10 values from a $\text{Bin}(10, 0.01)$ distribution, we can see that the majority of the mass of the generalized fiducial solution is below the Bayesian. These results suggest that the generalized fiducial solution performs better in the case of small data size and small p and a near-identical performance in all other cases.

Figure 7 examines the upper confidence intervals for the generalized fiducial and bayesian approaches of 1000 simulations of a size 100 draw from a $\text{Bin}(10, \cdot)$ distribution. The horizontal axis of Figure 7 is the coverage probability we set for a given credible interval based off of the posterior distribution we derive, while the vertical axis is the observed proportion of time that a given credible interval captured the true parameter.

Figure 7 reaffirms what we saw in Figures 5 & 6. Namely, that the bayesian solution is not significantly different from the generalized fiducial for values of p other than 0.01. We

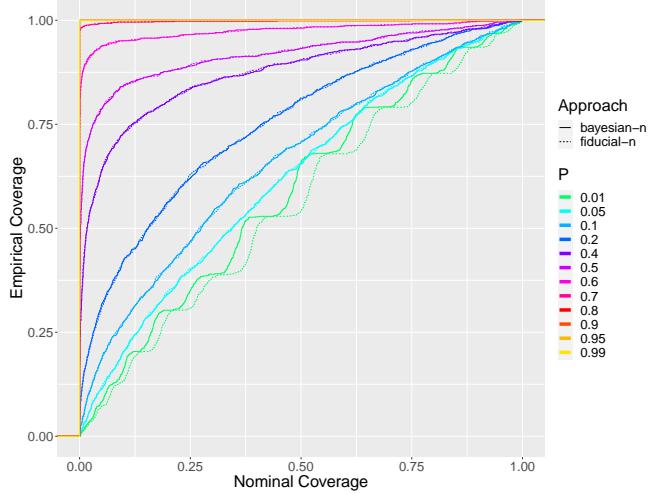


FIG 7. The coverage of credible intervals at varying levels of confidence for the generalized fiducial and bayesian approaches. This figure was calculated in the same was as Fig. 1.

also observe that for larger values of p an interval at any coverage level will contain the true parameter value. In the case of $p = 0.99$, a credible interval at any non-zero confidence level captures the true parameter with complete certainty. This relationship between nominal and empirical coverage for larger values of p follows from what we observe in Figure 5, which suggests that our posterior distribution is just a single point (the true parameter value n_0), and thus credible intervals of at any confidence level are simply the set $\{n_0\}$. For smaller values of p the coverage plots are closer to the 45 degree line indicating a close to correct coverage. The waves in the plot are caused by discretization effects.

Our results show that inference on n for known p is fairly straightforward when the probability of success is high. For smaller values of p , while both the generalized fiducial and bayesian approaches have less accuracy, they remain comparable. In the case of smaller values of p with a smaller number of observations on which to do inference, we can see a slightly better performance with the generalized fiducial approach.

3.3. Binomial Distribution; n unknown, p unknown. We will now consider our final non-trivial example through the fiducial lens: simultaneous inference on (n, p) for the binomial distribution. As far as we are aware, this problem has yet to receive a satisfactory solution using any philosophy whatsoever.

We consider a set of m independent binomial data: $\mathbf{Y} = (Y_1, \dots, Y_m)$ i.i.d. $\text{Bin}(n, p)$, where both p and n are unknown and our objective is inference on the pair (n, p) . We will use the same DGA (9) as in the previous example. Recall the following identity relating the binomial binomial CDF F and the beta CDF G such that $F_{n,p}(y) = G_{n-y+1,y}(1-p)$ (Wadsworth, 1960). This allows us to write an equivalent form of (9):

$$(10) \quad G_{Y_i+1,n-Y_i}^{-1}(1-U_i) \geq p > G_{Y_i,n-Y_i+1}^{-1}(1-U_i), \quad i = 1, \dots, m,$$

which we can invert to get

$$\hat{p}_{i,n}^{upper} = G_{Y_i+1,n-Y_i}^{-1}(1-U_i), \quad \hat{p}_{i,n}^{lower} = G_{Y_i,n-Y_i+1}^{-1}(1-U_i), \quad i = 1, \dots, m.$$

For any value \hat{n} , the set of associated values p that satisfy (10) is the interval

$$(\max\{\hat{p}_{1,\hat{n}}^{lower}, \dots, \hat{p}_{m,\hat{n}}^{lower}\}, \min\{\hat{p}_{1,\hat{n}}^{upper}, \dots, \hat{p}_{m,\hat{n}}^{upper}\}).$$

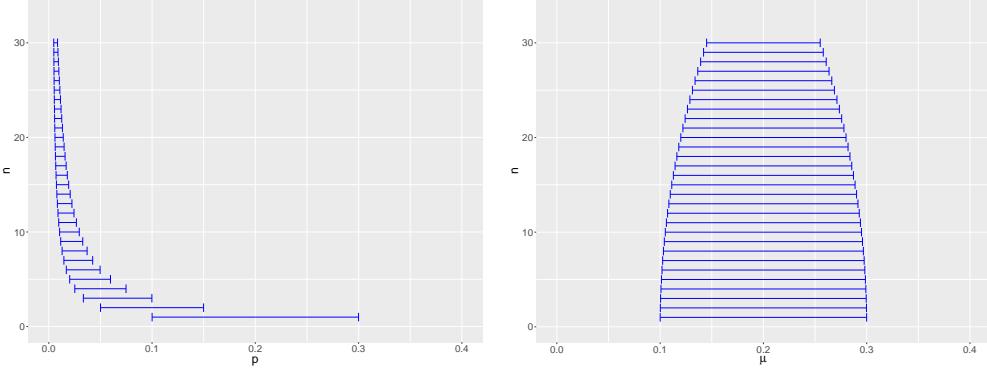


FIG 8. Two parameterizations for the same single draw from the GFD: using n, p parameterization (on the left) versus using n, μ parameterization (on the right). Note that a draw from this GFD defines a region of values in a joint parameter space that is discrete in n and continuous in p and μ .

The interval is taken as an empty set if the lower limit is larger than the upper limit.

A generalized fiducial solution to this problem would define a GFD that assigns probability mass to sets

$$\bigcup_{\hat{n}} \{\hat{n}\} \times (\max\{\hat{p}_{1,\hat{n}}^{lower}, \dots, \hat{p}_{m,\hat{n}}^{lower}\}, \min\{\hat{p}_{1,\hat{n}}^{upper}, \dots, \hat{p}_{m,\hat{n}}^{upper}\}).$$

These sets contain multiple predicted values \hat{n} that are each associated with a set of values of p . The left panel of Figure 8 shows a single random draw from the GFD on the parameter space (n, p) . Note that a single draw in the two-dimensional parameter space is a collection of horizontal segments at each \hat{n} . Naturally, larger and larger values of \hat{n} correspond with smaller values for \hat{p} , making them difficult to visualize. For this reason, and because of some desirable properties of the limiting distribution of $\hat{n}\hat{p}$ discussed in Appendix D, we propose to work with a reparameterization (n, μ) , where $\mu := np$. This results in a GFD that assigns mass to sets of the form

$$\bigcup_{\hat{n}} \{\hat{n}\} \times (\max\{\hat{\mu}_{1,\hat{n}}^{lower}, \dots, \hat{\mu}_{m,\hat{n}}^{lower}\}, \min\{\hat{\mu}_{1,\hat{n}}^{upper}, \dots, \hat{\mu}_{m,\hat{n}}^{upper}\}),$$

where

$$\hat{\mu}_{i,\hat{n}}^{upper} = \hat{p}_{i,\hat{n}}^{upper} \hat{n}, \quad \hat{\mu}_{i,\hat{n}}^{lower} = \hat{p}_{i,\hat{n}}^{lower} \hat{n}, \quad i = 1, \dots, m.$$

A draw from this reparameterized GFD can be seen in the right panel of Figure 8.

Our algorithm for simulating values from the GFD of (μ, n) uses a Metropolis within Gibbs sampler. We start with a set of values \mathbf{U} that give a solution that is not the empty set, we record this solution set, then we resample the \mathbf{U} values one by one such a way that the DGA (9) is still satisfied. These updates are available in closed form. To improve mixing we also add to each Gibbs scan two additional random walk Metropolis Hastings steps, one in the n direction and one in the μ direction, so that we can adequately explore our sample space. Further details on the MCMC algorithm can be found in Appendix D.

Figures 9 & 10 show four implementations of our Gibbs sampler for different true values of n and p . For these figures we draw 50 observations from the distributions $Bin(0.1, 15)$, $Bin(0.1, 75)$, $Bin(0.9, 15)$, and $Bin(0.9, 75)$, respectively.

For p small, the binomial distribution approaches a Poisson distribution and results in large sets of \hat{n} values for each draw from our GFD. One can observe this in Figure 9, where certain draws from the GFD seem to go upward to infinity in the n direction. Our algorithm senses

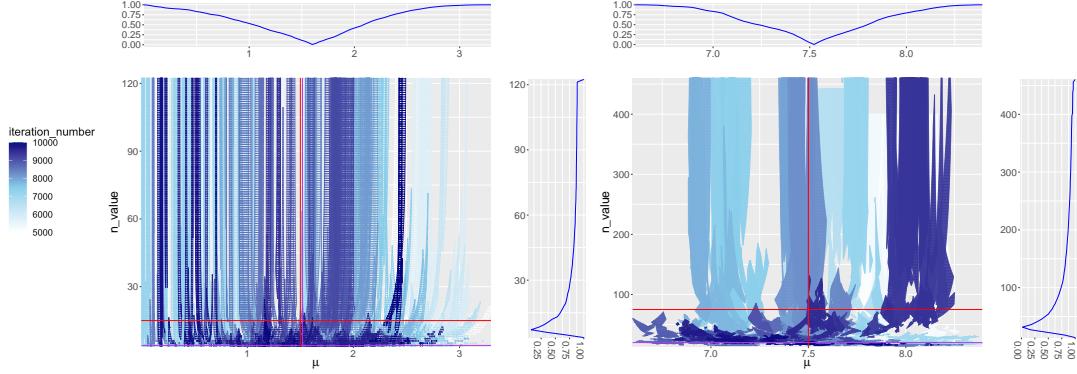


FIG 9. GFD for $\text{Bin}(15, 0.1)$ data (left) and $\text{Bin}(75, 0.1)$ data (right). Different shades of color in the center plot correspond to separate draws using a Gibbs sampler on the GFD built on the data. The top and right marginal plots show the confidence curves of μ and n values, respectively, centered at the median. The red lines cross at the true (n, μ) value used to generate the distribution and the purple lines cross at a value of DasGupta and Rubin (2004) estimator (\hat{n}, \bar{Y}) that are used to initialize our algorithm.

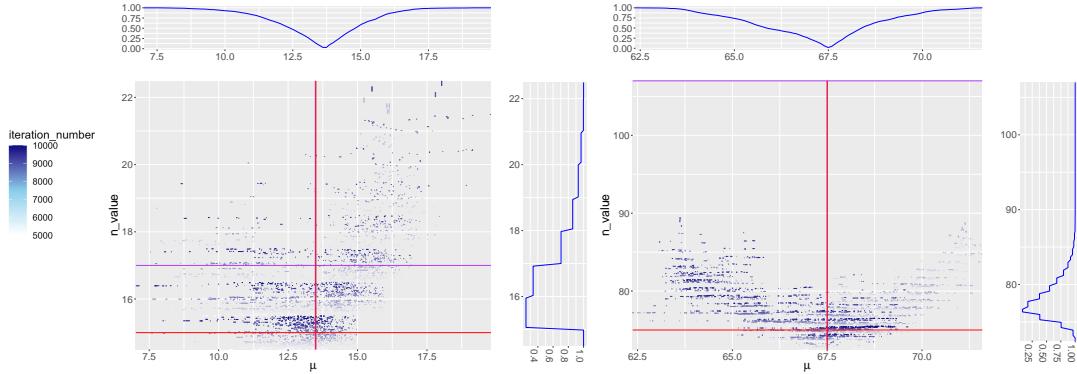


FIG 10. GFD for $\text{Bin}(15, 0.9)$ data (left) and $\text{Bin}(75, 0.9)$ (right) visualized in the same way as Figure 9.

when a draw from the GFD is reaching its limiting Poisson distribution and stops checking new values for n . The specific details of this stopping criterion are discussed in Appendix D. In contrast, for large p , the algorithm reports a narrower range in n (Figure 10). Note that for clarity we jittered the graph in the n direction, especially in the case of Figure 10, where most of the sets have a n value concentrated in one area.

Our Gibbs sampler is able to explore both the n space and the μ space simultaneously. We examine the marginal movement in the μ and the n directions in Figures 11 & 12. To create these figures, we select a single representative μ and n from each draw from the GFD. To select a single μ from a set of values in a single draw from the GFD, we randomly choose either the minimum or maximum μ value. We perform the same process in the n direction; for every draw from the GFD, we randomly select either the largest or smallest n value present to be that draw's representative in the trace plot. For each of these selections, the minimum or maximum are chosen with equal probability. The resulting figures show how our algorithm explores the n and μ space around the true parameter values. For smaller true values of p , these plots show how our algorithm visits very large values of n . These draws correspond to the instances where the algorithm begins to capture the limiting Poisson distribution.

To show how this approach tends towards the true parameter values, we generated data for $\text{Bin}(75, 0.9)$ data and started the algorithm at $n = 107$, a value much larger than the truth. Figures 13 & 14 shows how despite the extreme error in our initial guess, the Gibbs

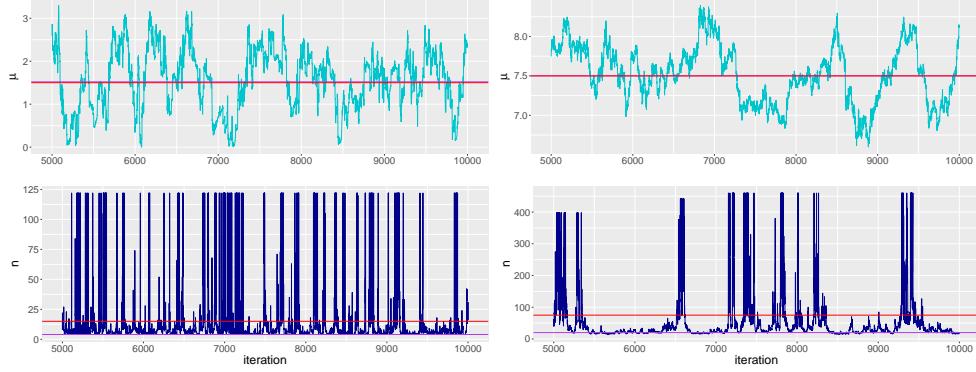


FIG 11. Marginal trace plots for μ and n for $\text{Bin}(15, 0.1)$ data (left) and $\text{Bin}(75, 0.1)$ data (right). These plots use the same GFD produced for figures 9 and 10. The red line crosses at the true parameter values used to generate the data and the purple line crosses at a value of [DasGupta and Rubin \(2004\)](#) estimator (\hat{n}, \bar{Y}) that are used to initialize our algorithm.

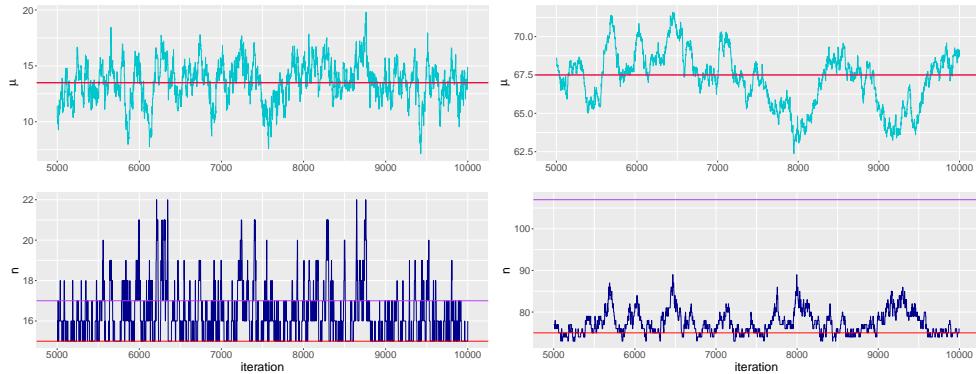


FIG 12. Marginal trace plots for μ and n for $\text{Bin}(15, 0.9)$ data (left) and $\text{Bin}(75, 0.9)$ (right) visualized in the same way as Figure 11.

sampler will eventually walk towards the true parameter values, which are marked by the red horizontal line.

We evaluate this method via a simulation study. For each combination $n \in \{15, 75\}$ and $p \in \{0.1, 0.5, 0.9\}$, we drew 50 observations from a $\text{Bin}(n, p)$ distribution 1000 times. One important thing to note about this solution is that each element of the GFD is a *set* of values, rather than a singular pair $(\hat{n}, \hat{\mu}_n)$. This complicates the notion of containment ratios slightly. Ideally, for each iteration of our simulation study, we would like to be able to calculate a box of reasonable values that we would expect to contain the true parameter pair (n, μ) 95% of the time. We followed the Dempster-Shafer theory from Bayesian analysis ([Dempster, 2008](#)) to draw Belief and Plausibility boxes, which we then used to calculate containment ratios. In this application, the Belief of our posterior was the box that fully contained 95% of our posterior draws, while the Plausibility of our posterior was the box that simply intersected 95% of our posterior draws. One can see the resulting containment ratios using Belief and Plausibility in Table 3. Something of note is that, in this application, Belief is often larger than Plausibility. This is the opposite of what is typical for Dempster-Shafer theory, but nonetheless accurate. A box that must contain 95% of our intervals in their entirety will likely be larger than a box that merely needs to “touch” 95% of them.

Table 3 shows that our method performs conservatively for smaller realizations of the parameter value n . Our coverage probabilities drop as both n and p become larger, and for

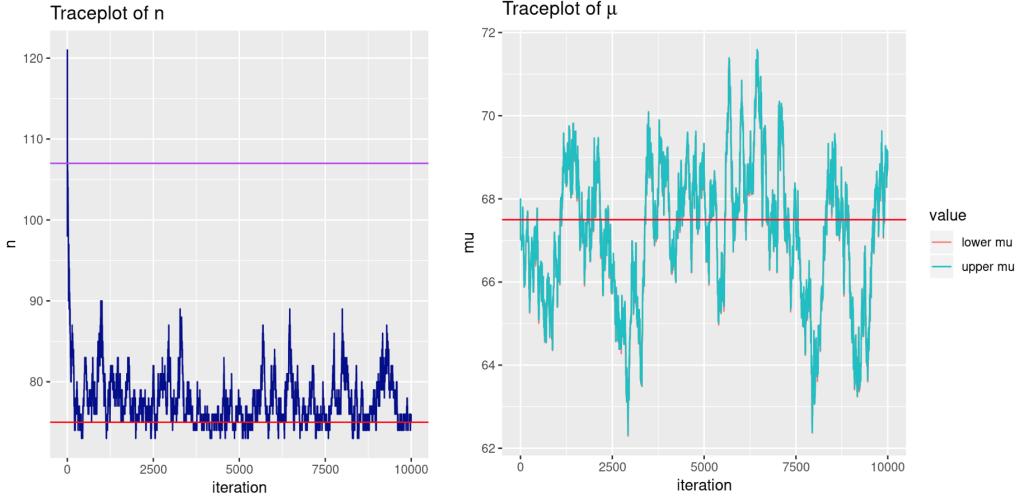


FIG 13. Traceplot of n for $\text{Bin}(0.9, 75)$ with a first proposed value of $\hat{n} = 107$

FIG 14. Traceplot of μ for $\text{Bin}(0.9, 75)$ with a first proposed value of $\hat{\mu} = 67.5$

Distribution	95% Plausibility Containment	95% Belief Containment
$\text{Bin}(0.1, 15)$	1	1
$\text{Bin}(0.1, 75)$	0.98	0.97
$\text{Bin}(0.5, 15)$	0.85	0.86
$\text{Bin}(0.5, 75)$	0.84	0.86
$\text{Bin}(0.9, 15)$	1	1
$\text{Bin}(0.9, 75)$	0.78	0.78

TABLE 3
The coverage of 95% Belief and Plausibility boxes on posterior fiducial sets of (n, μ) .

$p = 0.5$. We attribute the issues related to large n to the fact that our algorithm has a much larger space of candidate \hat{n} values to search as the true parameter value n increases. To counteract this, our MCMC chain would need to run for significantly longer to reach a steady state on the true parameter value.

4. Conclusion. GFI provides an alternative perspective on numerous classical inferential problems. Our selection of examples show how the generalized fiducial framework can derive a meaningful, practically feasible distribution on a target parameter without a need for an arbitrarily defined prior. While there is much to be learned from using this framework on classic theoretical problems, modern research has shown how the fiducial perspective has been helpful in understanding previously unsolved problems. We provide a short list of such references here for the interested reader: (E *et al.*, 2008; Cisewski and Hannig, 2008; Wandler and Hannig, 2011, 2012b,a, 2006; Hannig and Lee, 2009; Liu and Hannig, 2016).

5. Acknowledgements. Jan Hannig's research was supported in part by the National Science Foundation under Grant No. IIS-1633074 and DMS-1916115.

APPENDIX A: DERIVATION OF THE MULTIVARIATE JACOBIAN QUANTITY
 $J(\mathbf{y}; \mu, U, \Lambda)$ AND MARGINAL $r(\mathbf{A})$

Let $J^{i,j}$ be a matrix of all zeros save for a value of 1 at the index (i, j) . In general, the form of our Jacobian derivative with respect to any value of the skew-symmetric matrix A is $J(\mathbf{y}, \mu, A, \Lambda) = [\nabla_\mu \mathbf{Y}_i; \frac{\partial \mathbf{Y}_i}{\partial \lambda_j}; \frac{\partial \mathbf{Y}_i}{\partial a_{j,k}}]$, where “;” denotes the row concatenation of the following three quantities:

$$\begin{aligned}\nabla_\mu \mathbf{Y}_i &= I_d \\ \frac{\partial \mathbf{Y}_i}{\partial \lambda_j} &= \lambda_j^{-1}(I_d - A)(I_d + A)^{-1}J^{j,j}[(I_d - A)(I_d + A)^{-1}]^\top(\mathbf{Y}_i - \mu), \quad j = 1, \dots, d \\ \frac{\partial \mathbf{Y}_i}{\partial a_{j,k}} &= 2(I_d + A)^{-1}(-J^{i,j} + J^{j,i})(I_d - A)^{-1}(\mathbf{Y}_i - \mu), \quad 1 \leq j < k \leq d.\end{aligned}$$

Using the l_2 norm, the final form of our Jacobian is $D(J(\mathbf{y}, \mu, A, \Lambda))$ where $D(X) = |\det(\sqrt{X^T X})|$. One can show, using the Cauchy-Binet formula, that the μ vector drops out and that we can factor out the Λ matrix such that $D(J(\mathbf{y}, \mu, A, \Lambda)) = \det(\Lambda)^{-1}D^*(J(\mathbf{y}, A))$. We can thus write the GFD likelihood using (3) like so:

$$r_{\mathbf{y}}(A, \Lambda, \mu) \propto (2\pi)^{\frac{-nd}{2}} |\Sigma|^{-n/2} \exp \left\{ \frac{-1}{2} \text{tr} \left(\Sigma^{-1} \left(\sum_{i=1}^n (\mathbf{y}_i - \mu)(\mathbf{y}_i - \mu)^T \right) \right) \right\} D(J).$$

By integrating out μ and rearranging the terms that depend on λ_i we get

$$r_{\mathbf{y}}(A, \Lambda) \propto D^*(J(\mathbf{y}, A))(2\pi)^{\frac{-d(n-1)}{2}} n^{-d/2} \left[\prod_{i=1}^d \left(\frac{1}{\lambda_i^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-1}{2\lambda_i^2} (U^T SSE U)_{ii} \right\} \right],$$

Next, we integrate out Λ to get the marginal GFD of A

$$r_{\mathbf{y}}(A) \propto D^*(J(\mathbf{y}, A))(\pi)^{\frac{-d(n-1)}{2}} 2^{-d} n^{\frac{-d}{2}} \Gamma \left(\frac{n-1}{2} \right)^d \prod_{i=1}^d (U^T SSE U)_{ii}^{\frac{-(n-1)}{2}}.$$

A similar calculation can be done for an alternative data generating algorithm using minimal sufficient statistic $(\bar{\mathbf{Y}}, SSE)$. The only difference will be the form of $D^*(J(\mathbf{y}, A))$. See the STAN implementation for details.

APPENDIX B: DERIVATION OF THE MIXED MODELS JACOBIAN QUANTITY
 $J(\mathbf{Y}, S_\alpha, \mathbb{X}\beta, \sigma_\alpha, \sigma_e)$

The derivative $\frac{\delta \mathbf{Y}}{\delta \beta} = \mathbb{X}$ is trivial. To solve for the other two terms, note that $\Sigma = AA$ and therefore

$$\frac{\delta \Sigma}{\delta \sigma_\alpha^2} = A \frac{\delta A}{\delta \sigma_\alpha^2} + \frac{\delta A}{\delta \sigma_\alpha^2} A, \text{ where } \frac{\delta A}{\delta \sigma_\alpha^2} = \frac{1}{2} S_\alpha A^{-1},$$

The derivative $\frac{\delta A}{\delta \sigma_\alpha^2}$ commutes with A since

$$A = U \left(\text{diag} \left(\sqrt{\sigma_\alpha^2 \lambda_i + \sigma_e^2} \right) \right) U', \quad \frac{\delta A}{\delta \sigma_\alpha^2} = \frac{1}{2} U \left(\text{diag} \left(\lambda_i (\sigma_\alpha^2 \lambda_i + \sigma_e^2)^{-1/2} \right) \right) U',$$

and clearly $\frac{\delta A}{\delta \sigma_\alpha^2} A = A \frac{\delta A}{\delta \sigma_\alpha^2}$. Therefore,

$$\frac{\delta \mathbf{Y}}{\delta \sigma_\alpha^2} = \frac{\delta}{\delta \sigma_\alpha} (\mathbb{X}\beta + AZ) \Big|_{Z=A^{-1}(Y-\mathbb{X}\beta)} = \frac{1}{2} S_\alpha \Sigma^{-1} (\mathbf{Y} - \mathbb{X}\beta).$$

Following this same logic gives

$$\frac{\delta \mathbf{Y}}{\delta \sigma_e} = \Sigma^{-1}(\mathbf{Y} - \mathbb{X}\beta) = (\sigma_\alpha^2 S_\alpha + \sigma_e^2 I)^{-1}(\mathbf{Y} - \mathbb{X}\beta).$$

APPENDIX C: BINOMIAL ALGORITHM FOR UNKNOWN n , KNOWN p

As in Section 3.1, the generalized fiducial solution to the n unknown problem assigns mass to *sets* of integers rather than individual values. Our aim is to assign fiducial probabilities to all reasonable sets of n values in the sample space. Considering a *set* \mathbf{s} of candidate n values and fixing the observed data \mathbf{Y} , we start by calculating the probability that we observe a set of uniform values such that any $n^* \in \mathbf{s}$ could be the true n value. Notice that the set of viable uniform values for a superset \mathbf{s} must necessarily be a subset of the set of viable uniform values for s . The astute reader will notice that this is the same as the notion of *commonality* found in Dempster-Schaffer calculus (Shafer, 1976). Based on the following inversion of (9),

$$(11) \quad F_{n^*,p}(Y_i - 1) < U_i^* \leq F_{n^*,p}(Y_i), \quad i = 1, \dots, n.$$

Since $F_{n^*,p}(Y_i)$ is non-decreasing in n^* , we see that for any fixed (U_1, \dots, U_n^*) the set of n^* is an interval of consecutive integers. We will denote the set of all integer intervals $\mathbf{S} = \{\{i, i+1, \dots, j\}, 1 \leq i \leq j\}$. The commonality of the set $\mathbf{s} \in \mathbf{S}$ will be

$$p(\mathbf{s}|\mathbf{Y}) = \prod_{i=1}^m [F_{\max\{\mathbf{s}\},p}(y_i) - F_{\min\{\mathbf{s}\},p}(y_i - 1)]^+.$$

Our algorithm begins by using this commonality to address the issue of no strict upper bound on the set of possible values for n^* . First, select a precision cutoff ϵ . Then, we define our set of candidate n^* values, which we will call \mathbf{N} , by way of the following algorithm: Starting with the observed maximum of the data consider new values sequentially, comparing the commonality of the candidate n^* value: $p(\{n^*\}|\mathbf{Y})$ with commonalities observed so far. If the ratio of the commonality to the maximum commonality in \mathbf{N} accepted so far

$$\frac{p(\{n^*\}|\mathbf{Y})}{\max_{k \in \mathbf{N}} \{p(\{k\}|\mathbf{Y})\}} > \epsilon,$$

add n^* to our set \mathbf{N} and consider $n^* + 1$. Otherwise, halt the algorithm.

Next we approximate the GFD by restricting ourselves only to subsets of \mathbf{N} . Define $\hat{\mathbf{S}} = \{\mathbf{s} \in \mathbf{S}, \mathbf{s} \subset \mathbf{N}\}$. Using the analogue noted previously between Dempster-Shaffer calculus and the inversion of our DGA, we define the fiducial probability of a set \mathbf{s} as the Dempster-Shaffer *mass* of \mathbf{s} (Shafer, 1976; Yager and Liu, 2010) modified and renormalized so that $r_{\mathbf{Y}}(\emptyset) = 0$. We calculate the the Dempster-Shaffer mass of \mathbf{s} recursively by starting from $m(\mathbf{N}) = p(\mathbf{N}|\mathbf{Y})$ and then for $\mathbf{s} \in \hat{\mathbf{S}}$

$$m(\mathbf{s}) = \left[p(\mathbf{s}|\mathbf{Y}) - \left(\sum_{\mathbf{r} \in \hat{\mathbf{S}}: \mathbf{s} \subset \mathbf{r}} m(\mathbf{r}) \right) \right]^+.$$

Then, after the probability of all sets are defined in this way, we renormalize them so that they add to 1. The fiducial probability for each element of $\mathbf{s} \in \hat{\mathbf{S}}$ is

$$r_{\mathbf{Y}}(\mathbf{s}) = \frac{m(\mathbf{s})}{1 - m(\emptyset)}.$$

APPENDIX D: BINOMIAL ALGORITHM FOR UNKNOWN n , UNKNOWN p

We begin by making an important observation about the distribution of the predicted value \hat{p} . For random U_i^* , fixed \mathbf{Y} and n , we have that

$$\hat{p}_i^{upper} \sim Beta(Y_i + 1, n - Y_i), \quad \hat{p}_i^{lower} \sim Beta(Y_i, n - Y_i + 1).$$

Writing out the distribution on the bounds of our fiducial sets for p in this way allows use of the following well-known result:

LEMMA D.1. *Let $\mu := np$. Then*

$$\hat{\mu}_i^{upper} = n\hat{p}_i^{upper} \xrightarrow{\mathcal{D}} Gamma(Y_i + 1, 1) \quad \hat{\mu}_i^{lower} = n\hat{p}_i^{lower} \xrightarrow{\mathcal{D}} Gamma(Y_i, 1),$$

as $n \rightarrow \infty$.

PROOF. See, for instance, [Gut \(2005\)](#). ■

Our aim is to develop a method that can simulate draws from the GFD by calculating sets with elements of the form $\hat{n} \times (\hat{\mu}^{lower}, \hat{\mu}^{upper}) \in \mathbb{N} \times \mathcal{B}(\mathbb{R}_+)$. A simplistic approach to generating values from the GFD would be to repeatedly re-sample the set of random variables \mathbf{U} on the space $[0, 1]^m$ and in turn recalculate the paired sets that satisfy (9). There are two issues that arise for this approach. First, there is no upper bound on the set of possible n values, and second, it is possible that a majority of the uniform sets \mathbf{U} drawn completely randomly will yield no solution to (9). To address these issues we develop a stopping criterion for searching the sample space in the n direction and we develop a more computationally efficient way to randomly select uniform values \mathbf{U} .

We use the following protocol for choosing a range of potential n values. While there are many ways to choose a first candidate \hat{n} , we use the estimator suggested by [DasGupta and Rubin \(2004\)](#). From this first candidate we check each subsequent n by increment of one and calculate $(\hat{\mu}^{lower}, \hat{\mu}^{upper})$. We develop a stopping rule for \hat{n} based on the asymptotic result in Lemma D.1. Given a precision parameter $\epsilon > 0$, we stop looking for further \hat{n} values whenever

$$|\hat{\mu}_i^{upper} - H^{-1}(1 - U_i)_{Y_i+1,1}| < \epsilon \quad \text{and} \quad |\hat{\mu}_i^{lower} - H^{-1}(1 - U_i)_{Y_i,1}| < \epsilon,$$

for all $i \in \{1, \dots, m\}$, where $H_{\alpha, \beta}^{-1}$ is the quantile function of a $Gamma(\alpha, \beta)$ random variable. We stop searching for values of n that work as soon as we are sufficiently close to the limiting Gamma distribution, which does not depend on n . This stopping criterion leverages our assumption that our target data is binomial, and not Poisson, to reduce the space of possible n values to a finite set.

We implement an Gibbs sampler, sampling from the uniform distribution \mathbf{U} randomly selected in such a way that there is a new solution set that is non-empty, i.e., satisfying Equation (10), or equivalently

$$G_{n-Y_i, Y_i+1}(1-p) \geq U_i > G_{n-Y_i+1, Y_i}(1-p), \quad i = 1, \dots, m.$$

In particular, each value U_i is resampled iteratively by randomly selecting a new value from its conditional distribution given all the other U s; a uniform distribution on

$$\left(\min\{G_{n-Y_i+1, Y_i} \left(1 - \frac{\tilde{\mu}_{\tilde{n}, j}^{lower}}{\tilde{n}} \right) : j \neq i\}, \max\{G_{n-Y_i, Y_i+1} \left(1 - \frac{\tilde{\mu}_{\tilde{n}, j}^{lower}}{\tilde{n}} \right) : j \neq i\} \right],$$

where \tilde{n} , $\tilde{\mu}_{\tilde{n}, j}^{lower}$, and $\tilde{\mu}_{\tilde{n}, j}^{upper}$ for $j = \{1, \dots, m\} \setminus \{i\}$ are the solution set that we would get if we were to remove the i th observation from our data. To speed up the computation we

actually resample U_i s in batches, so that all U_i s corresponding to the same observed value of Y_i are sampled together.

Our investigations of the above method of resampling the uniform values \mathbf{U} have revealed that there is a need for steps that would result large changes to n and μ . To allow for this, we add two Metropolis-Hastings steps at the end of each Gibbs sampler scan, one in the n direction and one in the μ direction.

The MH step in the μ direction was implemented using the following algorithm. Assume that you already have finished the above protocol for choosing a new set of uniform values \mathbf{U} and that has a corresponding solution set

$$\bigcup_{\hat{n}} \{\hat{n}\} \times (\max\{\hat{\mu}_{1,\hat{n}}^{lower}, \dots, \hat{\mu}_{m,\hat{n}}^{lower}\}, \min\{\hat{\mu}_{1,\hat{n}}^{upper}, \dots, \hat{\mu}_{m,\hat{n}}^{upper}\}).$$

Randomly select a single value μ^* from the set of $\hat{\mu}$ values associated with the smallest \hat{n}

$$(\max\{\hat{\mu}_{1,\min\{\hat{n}\}}^{lower}, \dots, \hat{\mu}_{m,\min\{\hat{n}\}}^{lower}\}, \min\{\hat{\mu}_{1,\min\{\hat{n}\}}^{upper}, \dots, \hat{\mu}_{m,\min\{\hat{n}\}}^{upper}\}).$$

Then, draw a new μ^* from the proposal distribution $N(\mu^*, \sigma^2)$ for some predetermined σ . Calculate a set of uniform values, \mathbf{U}^* , such that μ^* is contained in the solution set. Let $\dot{n}, \{\dot{\mu}_{\min\{\dot{n}\},j}^{lower}\}_{j=1}^m$, and $\{\dot{\mu}_{\min\{\dot{n}\},j}^{upper}\}_{j=1}^m$ be the values in the solution set created by using Equation (10) with \mathbf{U}^* . The acceptance ratio for these new \mathbf{U}^* will then be

$$\frac{(\dot{\mu}_{\min\{\dot{n}\},j}^{lower} - \dot{\mu}_{\min\{\dot{n}\},j}^{upper}) \prod_{i=1}^m \left(G_{n-Y_i, Y_i+1} \left(1 - \frac{\hat{\mu}_{\dot{n},i}^{lower}}{\dot{n}} \right) - G_{n-Y_i+1, Y_i} \left(1 - \frac{\hat{\mu}_{\dot{n},i}^{lower}}{\dot{n}} \right) \right)}{(\dot{\mu}_{\min\{\dot{n}\},j}^{lower} - \dot{\mu}_{\min\{\dot{n}\},j}^{upper}) \prod_{i=1}^m \left(G_{n-Y_i, Y_i+1} \left(1 - \frac{\hat{\mu}_{\dot{n},i}^{lower}}{\dot{n}} \right) - G_{n-Y_i+1, Y_i} \left(1 - \frac{\hat{\mu}_{\dot{n},i}^{lower}}{\dot{n}} \right) \right)}.$$

The MH step in the n direction was implemented using the following algorithm. Like the MH step in the μ direction, start by randomly selecting a single μ^* value from set of $\hat{\mu}$ values associated with the smallest \hat{n} , $\mu^* \in (\hat{\mu}_{\min\{\hat{n}\},j}^{lower}, \hat{\mu}_{\min\{\hat{n}\},j}^{upper})$. Then, let $n^* := \min\{\hat{n}\} + 1 - 2X$, where $X \sim Bernoulli(1/2)$, be your new proposed n value. Next, find a set of uniform values \mathbf{U}^* such that n^* and μ^* are both in the solution set. Let $\dot{n}, \{\dot{\mu}_{\min\{\dot{n}\},j}^{lower}\}_{j=1}^m$, and $\{\dot{\mu}_{\min\{\dot{n}\},j}^{upper}\}_{j=1}^m$ be the values in the solution set created by using Equation (10) with \mathbf{U}^* . The acceptance ratio of \mathbf{U}^* has the same form as the acceptance ratio for the MH step in the μ direction.

The specific details of the algorithm can be read from our implementation (in R) on GitHub [<https://github.com/sirmurphalot/IntroductionGFI>]. On this page we have also posted a full pseudocode version of the algorithm for the curious reader.

REFERENCES

- Ahrens, H. and Pincus, R. (1981) On two measures of unbalancedness in a one-way model and their relation to efficiency. *Biometrical Journal*, **23**, 227–235. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.4710230302>.
- Berger, J. O., Sun, D. and Song, C. (2020a) Bayesian analysis of the covariance matrix of a multivariate normal distribution with a new class of priors. *Annals of Statistics*, **48**, 2381–2403.
- (2020b) An objective prior for hyperparameters in normal hierarchical models. *Journal of Multivariate Analysis*, 104606.
- Cisewski, J. and Hannig, J. (2008) Generalized fiducial inference for normal linear mixed models. *The Annals of Statistics*, **40**, 2102–2127.
- DasGupta, A. and Rubin, H. (2004) Estimation of binomial parameters when both n,p are unknown. *Journal of Statistical Planning and Inference*, **130**, 391–404.
- Dempster, A. (2008) The dempster-shafer calculus for statisticians. *International Journal of Approximate Reasoning*, **48**, 365 – 377. URL <http://www.sciencedirect.com/science/article/pii/S0888613X07000278>. In Memory of Philippe Smets (1938–2005).

- E, L., Hannig, J. and Iyer, H. (2008) Fiducial intervals for variance components in an unbalanced two-component normal mixed linear model. *Journal of the American Statistical Association*, **103**, 854–865. URL <https://doi.org/10.1198/016214508000000229>.
- Edlefsen, P., Liu, C. and Dempster, A. (2009) Estimating limits from poisson counting data using dempster-shafer analysis. *Annals of Applied Statistics*, **3**, 764–790.
- Eves, H. (1996) *Elementary Matrix Theory*. Allyn and Bacon, Inc., 1 edn. 265–267.
- Fazel, M., Hindi, H. and Boyd, S. P. (2003) Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proceedings of the 2003 American Control Conference, 2003.*, vol. 3, 2156–2162 vol.3.
- Fisher, R. (1935) The fiducial argument... *Annals of Eugenics*, **6**, 391–398.
- Förstner, W. and Moonen, B. (2003) *A Metric for Covariance Matrices*, 299–309. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Gut, A. (2005) *Probability: A Graduate Course*. Springer.
- Hannig, J. (2009) On Generalized Fiducial Inference. *Statistica Sinica*, **19**, 491–544.
- Hannig, J., E, L., Abdel-Karmin, A. and Iyer, H. (2003) Simultaneous fiducial generalized confidence intervals for ratios of means of lognormal distributions. *Austrian Journal of Statistics*, **35**, 261–269.
- (2007) Fiducial approach to uncertainty assessment: Accounting for error due to instrument resolution. *Metrologia*, **44**, 476–483.
- Hannig, J., Iyer, H., Lai, R. C. S. and Lee, T. C. M. (2016) Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, **111**, 1346–1361. URL <https://doi.org/10.1080/01621459.2016.1165102>.
- Hannig, J. and Lee, T. (2009) Generalized fiducial inference for wavelet regression. *Biometrika*, **96**, 847–860.
- Hannig, J., Wang, C. M. and Iyer, H. K. (2013) Uncertainty calculation for the ratio of dependent measurements. *Metrologia*, **4**, 177–186.
- Horn, R. A. and Johnson, C. R. (2012) *Matrix Analysis*. USA: Cambridge University Press, 2nd edn.
- Iyer, H., Wang, J. C. and Mathew, T. (2004) Models and confidence intervals for true values in interlaboratory trials. *Journal of the American Statistical Association*, **99**, 1060–1071.
- Konno, Y. (1995) Estimation of a normal covariance matrix with incomplete data under stein's loss. *Journal of Multivariate Analysis*, **52**, 308 – 324. URL <http://www.sciencedirect.com/science/article/pii/S0047259X85710160>.
- Liu, Y. and Hannig, J. (2016) Generalized fiducial inference for binary logistic item response models. *Psychometrika*, **81**, 290–324.
- McNally, R., Iyer, H. and Mathew, T. (2003) Tests for individual and population bioequivalence based on generalized p-values. *Statistics in Medicine*, **22**, 31–53.
- Neupert, S. D., Gowney, C. M., Zhu, X., Sorensen, J. K., Smith, E. L. and Hannig, J. (2020) Bff: Bayesian, fiducial, and frequentist analysis of cognitive engagement among cognitively impaired older adults. Submitted for publication.
- O'Dorney, E. (2014) Minimizing the cayley transform of an orthogonal matrix by multiplying by signature matrices. *Linear Algebra and its Applications*, **448**, 97 – 103. URL <http://www.sciencedirect.com/science/article/pii/S0024379514000615>.
- Schweder, T. and Hjort, N. L. (2016) *Confidence, likelihood, probability*, vol. 41. Cambridge University Press.
- Shafer, G. (1976) *A mathematical theory of evidence*. Princeton, New Jersey: Princeton University Press.
- Shi, W. J., Hannig, J., Lai, W. J. and Lee, T. C. M. (2017) Covariance estimation via fiducial inference. URL <http://arxiv.org/abs/1708.04929>. ArXiv:1708.04929.
- Stan Development Team (2020) RStan: the R interface to Stan. URL <http://mc-stan.org/>. R package version 2.21.2.
- Wadsworth, G. P. (1960) *Introduction to probability and random variables*. New York: McGraw-Hill, 1 edn. P. 52.
- Wandler, D. and Hannig, J. (2012a) Generalized fiducial confidence intervals for extremes. *Extremes*, **15**, 67–87.
- Wandler, D. V. and Hannig, J. (2006) Construction of fiducial confidence intervals for the mixture of cauchy and normal distributions. *Master's Thesis, Department of Statistics, Colorado State University*.
- (2011) Fiducial inference on the maximum mean of a multivariate normal distribution. *Journal of Multivariate Analysis*, **102**, 87–104.
- (2012b) A fiducial approach to multiple comparisons. *Journal of Statistical Planning and Inference*, **142**, 878–895.
- Wang, C. M. and Iyer, H. K. (2006a) Propagation of uncertainties in measurements using generalized inference. *Metrologia*, **42**, 145–153.
- (2006b) Uncertainty analysis for vector measurands using fiducial inference. *Metrologia*, **43**, 486–494.
- Wang, J., Hannig, J. and Iyer, H. (2012) Pivotal methods in the propagation of distributions. *Metrologia*, **49**, 382–389.

- Williams, J. P. and Hannig, J. (2019) Nonpenalized variable selection in high-dimensional linear model settings via generalized fiducial inference. *The Annals of Statistics*, **47**, 3, 1723–1753.
- Williams, J. P., Xie, Y. and Hannig, J. (2019) The eas approach for graphical selection consistency in vector autoregression models. *arXiv preprint arXiv:1906.04812*.
- Yager, R. R. and Liu, L. (2010) *Classic Works of the Dempster-Shafer Theory of Belief Functions*. Springer Publishing Company, Incorporated, 1st edn.
- Yang, R. and Berger, J. O. (1994) Estimation of a Covariance Matrix Using the Reference Prior. *The Annals of Statistics*, **22**, 1195–1211.