A Differential Geometric Perspective on Generalized Fiducial Inference

Yang Liu, Jan Hannig, Alexander C. Murph

October 12, 2022

Abstract

Generalized fiducial inference (GFI) produces Bayesian-like, post-data probabilistic
statements without specifying a prior distribution for model parameters. In the current
article, we propose a new characterization of the generalized fiducial distribution (GFD)
with the aid of differential geometry. Under suitable smoothness conditions, we establish
that the GFD has an absolutely continuous density with respect to the intrinsic measure of
a smooth manifold that is uniquely determined by the data generating equation. The
geometric analysis also sheds light on the connection and distinction between GFI and
Bayesian inference. Compared to the explicit expression of the fiducial density given by
Theorem 1 of Hannig et al. (2016), our new results can be applied to a broader class of
statistical models, including those with additional random effects. Furthermore, Monte
Carlo approximations to the GFD can be conveniently constructed via manifold Markov
chain Monte Carlo samplers. We provide an illustration using repeated-measures analysis
of variance.

Keywords: Generalized fiducial inference, Bayesian inference, approximate
Bayesian computation, differentiable manifold, Markov chain Monte Carlo
1 Introduction

Fiducial inference stemmed from the early work of [Fisher (1925, 1930, 1933, 1935)] and regained attention in recent years as an alternative inferential framework when little a priori information about model parameters can be garnered. Direct and indirect descendants of fiducial inference largely fall under two categories, focusing respectively on pre-data (i.e., frequentist) and post-data (i.e., conditional on the observed data) interpretation of the constructed data-dependent distribution of parameters. Methods such as probability matching priors [Fraser et al. 2010, Mukerjee and Dey, 1993, Mukerjee and Ghosh, 1997, Welch and Peers, 1963] and confidence distributions [Schweder and Hjort, 2002, Singh et al., 2005, Xie and Singh, 2013, Xie et al., 2011] belong to the former category: Probabilistic statements about parameters under the resulting data-dependent distribution are also (approximately) correct in the frequentist sense. Meanwhile, post-data probabilities can be loosely conceived as “belief” or “plausibility” for assertions on model parameters given the observed data. In the absence of prior beliefs, we can still obtain post-data inference based on the hypothesized data generating model: Exemplary methods include but are not limited to Dempster-Shafer theory [Dempster, 1966, 1968, 2008], inferential models [Martin and Liu, 2013, 2015a,c,b], and generalized fiducial inference (GFI; Hannig, 2009, 2013, Hannig et al, 2016). The current paper focuses on GFI and the post-data interpretation; unless otherwise specified, the observed data are treated as fixed in the sequel.

A statistical model specifies how data are generated through a data generating equation (DGE), which is a function of parameters and random components with completely known distributions (e.g., uniform or standard Gaussian variates). The DGE plays a key role in post-data statistical inference through the machinery of Monte Carlo simulation [Cranmer et al, 2020]. When a proper prior can be specified, we may simulate parameters and random components independently, obtain imputed data through the DGE, and retain the samples if and only if the imputed and observed data are “sufficiently

\footnote{A data generating equation may be referred to as a data generating algorithm (DGA; Murph et al., 2022a) when the generative process rather than the formal mathematical expression is of interest.}
close.” Such an accept-reject scheme is often referred to as approximate Bayesian computation (ABC; Beaumont, 2019; Beaumont et al., 2002): The retained samples of parameters approximately follow the posterior distribution and hence can be utilized to estimate posterior expectations. If no prior distribution is available, we can still sample random components but not parameters. To circumvent the latter, GFI proceeds to pair each realization of random components with the “optimal” parameter values such that the resulting imputed data is “as close to the observed data as possible.” Indeed such a “best matching” to the observed data may still not be “good enough”: Those values are deemed incompatible with the observed data and therefore have to be discarded, leading to a rejection step similar to ABC. It turns out that the resulting marginal samples of parameters (approximately) follow the *generalized fiducial distribution* (GFD; Hannig et al., 2016). The above heuristics will be made rigorous in later sections.

Although GFI has been successfully applied to many statistical problems such as linear mixed effects models (Cisewski and Hannig, 2012), high-dimensional regression (Lai et al., 2015), binary and ordinal logistic response models (Liu and Hannig, 2016, 2017), and covariance estimation (Shi et al., 2021), it remains difficult in general to design efficient algorithms to sample from the fiducial distribution as pointwise evaluation of the fiducial density is not always viable. As the most up-to-date result along this line, Hannig et al. (2016, Theorem 1) derived a closed-form expression of the fiducial density assuming that the random components and observed data have the same dimensionality (in addition to several other regularity conditions). However, their result does not apply to models with additional random effects, e.g., mixed-effects models.

In the present paper, we focus on the scenario where the DGE implicitly defines a smooth submanifold of the joint (Euclidean) space of parameters and random components, referred to as the data generating manifold. This setup applies broadly to common parametric models for continuous data, including random-effects models. As our main result, we establish that GFD is the parameter-marginal of an absolutely continuous distribution supported on the data generating manifold (Proposition 5). Because a
Bayesian posterior is subject to a similar characterization (Proposition 4), our geometric formulation reveals the fundamental differences between Bayesian inference and GFI. Moreover, it is possible to construct manifold Markov chain Monte Carlo (MCMC) samplers (e.g., Brubaker et al., 2012; Lelièvre et al., 2019, 2022; Zappa et al., 2018) targeting at distributions defined on the data generating manifold. When applied to GFI, the manifold MCMC s potentially scale up better than existing computational procedures for GFI.

The rest of the paper is organized as follows. We revisit in Section 2 the formal definitions of ABC and GFI; a graphical illustration is provided using a Gaussian location example. In Section 3 we first present a general result (Theorem 1): When an “ambient distribution” is truncated to a “sufficiently regular” sequence of increasingly finer approximations to a smooth manifold, the weak limit is absolutely continuous with respect to the manifold’s intrinsic measure. We then apply the general result to derive representations for Bayesian posteriors and GFDs (Propositions 4 and 5) and comment on their similarities and differences. We review in Section 4 an MCMC algorithm that (approximately) samples from distributions on differentiable manifolds. A repeated-measures analysis of variance (ANOVA) example is then presented to illustrate the sampling procedure (Section 5). Limitations and possible extensions of the proposed method are discussed at the end (Section 6).

2 Approximating Bayesian and Fiducial Inference by Simulation

2.1 Data Generating Equation

Let \( \mathcal{Y}, \mathcal{Y}, \) and \( \Theta \) denote the spaces of data, random components, and parameters associated with a fixed family of parametric models: In particular, \( \mathcal{Y} \subseteq \mathcal{R}^n, \mathcal{Y} \subseteq \mathcal{R}^m, \) and \( \Theta \subseteq \mathcal{R}^q, \) where \( n, m, \) and \( q \) are positive integers. Following Hannig et al. (2016), we characterize the model of interest by its DGE

\[
Y = G(U, \theta),
\]
in which the random components \( U \in \mathcal{Y} \) follow a completely known distribution (typically uniform or standard Gaussian), \( \theta \in \Theta \) denotes the parameters, and \( Y \in \mathcal{Y} \) denotes the random data. (1) can be conceived as a formalization of the data generating code: Given true parameters \( \theta \) and an instance of random components \( U = u \), a unique set of data \( Y = y \) can be imputed by evaluating the DGE, i.e., \( y = G(u, \theta) \).

Now suppose that we have observed \( Y = y \). Post-data inference aims to assign probabilities to assertions about parameters \( \theta \) conditional on the observed data \( y \) (Martin and Liu, 2015b). In the conventional Bayesian framework, we presume that \( \theta \) follows a proper prior distribution and make probabilistic statements based on the conditional distribution of \( \theta \) given \( y \). When no informative prior can be specified, one can still rely on objective priors that reflect the state of “lacking knowledge or information” (Kass and Wasserman, 1996; Berger, 2006; Berger et al., 2015). We next revisit the definition of a Bayesian posterior through the lens of ABC, as well as Hannig et al.’s (2016) definition of GFD: The latter replaces the prior sampling of parameters in ABC by an optimization problem in the parameter space, which is a natural workaround when no prior information is available.

2.2 Approximate Bayesian Computation

Let \( \rho \) denote the density of \( U \), and \( \pi \) be the prior density of \( \theta \); we only restrict to density functions with respect to the Lebesgue measure and assume that random number generation from \( \rho \) and \( \pi \) is feasible. Given the observed data \( y \) and a pre-specified tolerance level \( \varepsilon > 0 \), ABC is a simulation-based inferential procedure that repeatedly executes the following steps:

1) sample \( U \sim \rho \);
2) sample \( \theta \sim \pi \) independent of \( U \);
3) accept the draws if \( \|G(U, \theta) - y\| \leq \varepsilon \) and otherwise reject.

\( ^2 \) The model of interest characterized by the DGE \( G(\cdot, \cdot) \) need not coincide with the \( y \)-generating model. In fact, we do not postulate any specific generating mechanism for the observed data and only focus on post-data inference for the model of interest.
The above accept-reject sampling scheme constructs a truncated distribution on $\Upsilon \times \Theta$ with the following density:

$$\pi_\varepsilon(u, \theta|y) \propto \pi(\theta) \rho(u) \mathbb{I}_{\|G(u, \theta) - y\| \leq \varepsilon}(u, \theta),$$

in which $\| \cdot \|$ denotes the $\ell_2$-norm on the data space $\Upsilon$, and $\mathbb{I}_A$ denotes the indicator function for a set $A$. Integrating out $u$ results in

$$\pi_\varepsilon(\theta|y) \propto \pi(\theta) \mathbb{P}\{\|G(U, \theta) - y\| \leq \varepsilon|\theta\}.$$  (3)

Suppose that $Y$ has an absolutely continuous density $f(y|\theta)$ with respect to the Lebesgue measure on $\Upsilon$. (3) approximates the posterior

$$\pi(\theta|y) \propto \pi(\theta) f(y|\theta)$$

in the sense that, given $\theta$ and a small $\varepsilon$,

$$\mathbb{P}\{\|G(U, \theta) - y\| \leq \varepsilon|\theta\} \approx \lambda_Y \{y' \in \Upsilon : \|y' - y\| \leq \varepsilon\} f(y|\theta) \propto f(y|\theta)$$

when viewed as a function of $\theta$, in which $\lambda_Y$ denotes the Lebesgue measure on the data space. Readers are referred to Marin et al. (2012) and Beaumont (2019) for more comprehensive surveys of ABC.

### 2.3 Generalized Fiducial Inference

When prior information about $\theta$ is absent, we can no longer sample $\theta \sim \pi$ in the second step of the ABC recipe. Nevertheless, we are still able to determine whether the imputed random component $U$ can possibly reproduce the observed data $y$ (up to the pre-specified tolerance $\varepsilon$). Let

$$\hat{\theta}(y, U) = \arg \min_{\theta \in \Theta} \|G(U, \theta) - y\|.$$  (4)
The rationale of GFI is to pair each $U$ with the parameter values $\hat{\theta}(y,U)$ such that $G(U,\hat{\theta}(y,U))$ gives the closest approximation to $y$. ABC can then be modified into a Monte Carlo recipe for (approximate) GFI once we replace the prior sampling step by setting $\theta$ to $\hat{\theta}(y,U)$ and leave everything else intact. This modified procedure simulates from the truncated distribution

$$U \mid \left\{ U : \|G(U,\hat{\theta}(y,U)) - y\| \leq \varepsilon \right\}$$

(5)
on $\Upsilon$, which further induces a distribution on $\Theta$ via the map $\hat{\theta}(y,\cdot)$.

Hannig et al. (2016) went one step further and defined the GFD as the weak limit

$$\lim_{\varepsilon \downarrow 0} \left[ \hat{\theta}(y,U) \mid \left\{ U : \|G(U,\hat{\theta}(y,U)) - y\| \leq \varepsilon \right\} \right]$$

(6)

provided the limit exists. Assuming $n = m$ and several regularity conditions on the DGE (Assumptions A.1–A.4), Hannig et al. (2016) showed that the fiducial density corresponding to (6) can be expressed as

$$\psi(\theta|y) \propto \det \left( \nabla_{\theta}G(\hat{u}(y,\theta),\theta)^\top \nabla_{\theta}G(\hat{u}(y,\theta),\theta) \right)^{1/2} f(y|\theta),$$

(7)
in which $\hat{u}(y,\theta) \in \Upsilon$ satisfies $y = G(\hat{u}(y,\theta),\theta)$, and $\nabla_{\theta}G(u,\theta)$ denotes the $n \times q$ Jacobian matrix of $G(u,\theta)$ with respect to $\theta$. Hannig et al. (2016) also considered substituting the $\ell_\infty$- and $\ell_1$-norm for the $\ell_2$-norm in (6), which in general leads to fiducial densities different from (7) when $n > q$. We, however, focuses on the $\ell_2$ case in the present paper.

2.4 An Illustrative Example

Consider the Gaussian location model $Y \sim N(\mu,1)$ with the mean parameter $\mu \in \mathcal{R}$. For ease of graphical display, we focus on the transformed parameter $\theta = \Phi(\mu) \in (0,1)$, where $\Phi(\cdot)$ denotes the distribution function of $N(0,1)$. We express the

---

3 The minimum is assumed to exist for the moment; in case the minimum is not unique, an arbitrary one is selected (potentially at random).

4 The assumed regularity conditions guarantee that $\hat{u}(y,\theta)$ uniquely exists, and that the Jacobian matrix is defined and of full column rank.
corresponding DGE as

\[ Y = \Phi^{-1}(U) + \Phi^{-1}(\theta), \]

in which \( U \sim \text{Unif}[0, 1] \), and \( \Phi^{-1} \) is the inverse of \( \Phi \) (i.e., the standard Gaussian quantile function). The observed data \( y \) value is fixed at \(-0.5\).

For Bayesian inference, suppose that \( \theta \) follows a \( \text{Unif}[0, 1] \) prior, which implies a \( \mathcal{N}(0, 1) \) prior for the mean \( \mu \). It is straightforward to verify that the posterior density is

\[ \pi(\theta|y) = \phi\left(\Phi^{-1}(\theta)\right)^{-1} \sqrt{2} \phi\left(\sqrt{2}(\Phi^{-1}(\theta) - y/2)\right), \]

(9)

where \( \phi(\cdot) \) stands for the standard Gaussian density. Following the ABC recipe, we simulated \( U \) and \( \theta \) independently from \( \text{Unif}[0, 1] \), shown as evenly scattered dots over \( \Upsilon \times \Theta = [0, 1]^2 \) on the left panel of Figure 1. With the tolerance \( \varepsilon = 0.05 \), only \((u, \theta)\) pairs that satisfy \(|\Phi^{-1}(u) + \Phi^{-1}(\theta) - (-0.5)| \leq 0.05\) (dark gray colored dots) survive in the accept-reject step. The empirical \( \theta \)-marginal distribution of the retained draws closely resembles (9).

Meanwhile, the fiducial density (7) reduces to

\[ \psi(\theta|y) = \phi\left(\Phi^{-1}(\theta)\right)^{-1} \phi\left(x - \Phi^{-1}(\theta)\right). \]

(10)

Except for \( u \in \{0, 1\} \), which has probability zero under \( \text{Unif}[0, 1] \), \( \hat{\theta}(-0.5, u) = \Phi(-0.5 - \Phi^{-1}(u)) \) ensures \(|\Phi^{-1}(u) + \Phi^{-1}(\hat{\theta}(y, u)) - (-0.5)| = 0\). Essentially all the imputed \( u \)'s are therefore retained regardless of the value of \( \varepsilon \) in the simulation-based fiducial recipe. We associate each \( u \) with \( \theta = \hat{\theta}(-0.5, u) \) and plot \((u, \theta)\)'s on the right panel of Figure 1. It is observed that the \( u \)-marginal distribution remains uniform, and (10) can be well approximated by the histogram of \( \theta \).

We learn from the aforementioned illustration that, on the joint space of \( u \) and \( \theta \),

---

5 The normalizing constant is 1.
Figure 1: Graphic illustration of the Gaussian location example with $y = -0.5$. Left: approximate Bayesian computation. Samples of random components ($u$) and parameters ($\theta$) are represented as light gray dots in the unit square. ($u, \theta$) pairs that are sufficiently close to the curve $y = \Phi^{-1}(u) + \Phi^{-1}(\theta)$ are kept and highlighted in dark gray (acceptance rate = 2.64%). The empirical marginal distributions of the retained samples are displayed as histograms, with the theoretical posterior superimposed on the $\theta$-marginal. Right: fiducial inference. 100% of the imputed $u$’s are accepted, and each $u$ is paired with $\theta = \Phi(y - \Phi^{-1}(u))$. The empirical marginal distributions of the retained samples are displayed as histograms, with the theoretical fiducial density superimposed on the $\theta$-marginal. Dots fall exactly on the curve but are slightly jittered for clearer visualization.

Simulation-based Bayesian and fiducial inferences produce distributions that concentrate on

$$\mathcal{G}(y) = \{(u, \theta)^\top \in (0, 1)^2 : \Phi^{-1}(u) + \Phi^{-1}(\theta) = y\}$$

(11)

as $\varepsilon \downarrow 0$. The closure of (11), i.e., $\bar{\mathcal{G}}(y) = \mathcal{G}(y) \cup \{(0, 1), (1, 0)\}$, is a one-dimensional smooth manifold (with boundary) embedded in the unit square (highlighted as the black solid line in Figure 1). Similar characterizations can be established in a broader class of statistical models for continuous data, which we explicate in the next section.

3 Geometry of Statistical Inference

We have seen in our previous discussion that both the accept-reject ABC and the simulation-based fiducial recipe involve restricting ambient distributions to regions whose
sizes are controlled by $\varepsilon$ (see (2) and (5) for details). We pay heed to the special case that the regions of truncation contract to a smooth submanifold as $\varepsilon \downarrow 0$.

### 3.1 A General Weak Convergence Result

Our first result is completely general: It concerns the weak convergence of a sequence of truncated distributions to a limit that is supported on an implicitly defined submanifold. Let $h(x)$ be a map from $\mathcal{R}^d$ to $\mathcal{R}^n$ where $d > n$, $\mathcal{M}_\varepsilon = \{x \in \mathcal{R}^d : \|h(x)\| \leq \varepsilon\}$ where $\varepsilon > 0$, and $a : \mathcal{R}^d \rightarrow [0, \infty)$ be a non-negative function such that $0 < \int_{\mathcal{M}_\varepsilon} a(x)dx < \infty$ for all $\varepsilon > 0$. Although $a$ is not necessarily integrable over the entire ambient space $\mathcal{R}^d$, it is referred to as an *ambient density* in the sequel: Integrable and non-integrable $a$'s are respectively termed as *improper* and *proper* densities. Let $P_\varepsilon$ be the probability measure determined by the density $a(x)\mathbb{I}_{\mathcal{M}_\varepsilon}(x) / \int_{\mathcal{M}_\varepsilon} a(x)dx$. Under suitable assumptions, Theorem 1 characterizes the weak limit of $P_\varepsilon$ as $\varepsilon \downarrow 0$ by a density function with respect to $\lambda_\mathcal{M}$—the Riemannian measure on the submanifold

$$\mathcal{M} = \{x \in \mathcal{R}^d : h(x) = 0\}$$
induced by the Euclidean metric on the ambient space $\mathcal{R}^d$ (Lee, 2013, Chapter 13). The proof can be found in Appendix A.

**Theorem 1.** Suppose that

(i) $h$ is a twice continuously differentiable submersion;

(ii) $a$ is continuous and $\lambda_{\mathcal{M}}(\text{supp}(a) \cap \mathcal{M}) > 0$.

(iii) the collection of probability measures $\{P_\varepsilon : \varepsilon > 0\}$ is tight;

Then $P_\varepsilon \Rightarrow P_0$ as $\varepsilon \downarrow 0$, where $P_0$ has the following absolutely continuous density with respect to $\lambda_\mathcal{M}$:

$$f(x) = \frac{a(x) \det (\nabla h(x)\nabla h(x)^\top)^{-1/2}}{\int a(x') \det (\nabla h(x')\nabla h(x')^\top)^{-1/2} \lambda_\mathcal{M}(dx')}$$

for almost every $x \in \mathcal{M}$.

**Remark 1.** For a smooth submanifold $\mathcal{M} \subset \mathcal{R}^d$, we are guaranteed to have a collection of smooth local diffeomorphisms—called *coordinate charts*—that map elements of an open cover of $\mathcal{M}$ to subsets of a lower-dimensional Euclidean space. Local integration over a set on $\mathcal{M}$ fixes a coordinate chart that covers this set and performs the usual Lebesgue
integration on the Euclidean space pushed forward by the diffeomorphism: The inverse of the diffeomorphic map is often referred as a local parameterization of \( \mathcal{M} \). The rescaling for this change of variables depends on the Riemannian metric, which determines a Riemannian measure on the manifold that is invariant to the choice of the coordinate chart. Local integration is extended to global integration via a smooth partition of unity: a collection of weighting functions that allow one piece together the local integrals. Precise definitions of the aforementioned terms can be found in, e.g., Lee (2013).

**Remark 2.** Theorem 1 is inspired by Theorem 3.1 of Hwang (1980). Hwang’s result was proved for a sequence of Gibbs measures that concentrate on the minimum of an energy function. The collection of minimum energy states, or equivalently the limiting manifold, is required to be compact, which is restrictive but often suffices for optimization purposes in statistical physics. In contrast, our result applies to sequentially restricting a known ambient distribution to finer approximations of the data generating manifold—i.e., sublevel sets of \( h \), which is often unbounded for parametric statistical models.

**Remark 3.** Assumption (iii), i.e., the tightness of the measures \( \{P_\varepsilon\} \), automatically holds if \( \mathcal{M}_\varepsilon \) is compact for sufficiently small \( \varepsilon \)'s. When all the sublevel sets of \( h \) are unbounded, however, whether or not \( \{P_\varepsilon\} \) is tight is determined by the “tail behavior” of \( a \) and \( h \). Notably, \( a \) being a proper ambient density alone does not guarantee tightness. The following example in a two-dimensional ambient space demonstrates that \( \{P_\varepsilon\} \) can still be tight when \( a \) is improper but the sublevel set of \( h \) tapers off quickly on the tail, and that \( \{P_\varepsilon\} \) may not be tight when \( a \) is proper but the sublevel set of \( h \) has an exploding tail.

**Example 1.** Let \( x = (x_1, x_2)^\top \in [0, \infty)^2 \) and consider the constraint function

\[
h(x) = \frac{x_2}{g(x_1)},
\]

in which \( g \) is a positive function. The resulting \( \varepsilon \)-enlarged data generating set is

\[
\mathcal{M}_\varepsilon = \{ x \in [0, \infty)^2 : x_2 \leq \varepsilon g(x_1) \}.
\]
Figure 2: Sublevel sets $M_\varepsilon = \{ x \in [0, \infty)^2 : x_2 \leq \varepsilon g(x_1) \}$ in Example 1, where $g$ is a positive function. Left: $g(x_1) = \exp(-x_1^2/2)$, which vanishes quickly as $x_1 \to \infty$. Right: $g(x_1) = x$, which grows to infinity as $x_1 \to \infty$. The gray and shaded regions correspond to the sublevel sets when $\varepsilon = 0.05$ and $0.01$, respectively.

As $\varepsilon \downarrow 0$, $M_\varepsilon \downarrow M_0 = \{ x \in [0, \infty)^2 : x_2 = 0 \}$.

We first set $a(x) \equiv 1$ and $g(x_1) = \exp(-x_1^2/2)$ (left panel of Figure 2). Even though $a(x)$ is not integrable on the ambient space $[0, \infty)^2$, $g(x_1)$ is integrable on $[0, \infty)$. Hence, $a(x) I_{M_\varepsilon} a(x) dx$ is a valid density function that defines the probability measure $P_\varepsilon$. Consider the compact set $K = [0, C] \times [0, 1]$, in which $C > 0$. For all $\varepsilon < 1$,

$$P_\varepsilon \{ K \} = \frac{\int_0^C \left[ \int_0^{\varepsilon \exp(-x_1^2/2)} dx_2 \right] dx_1}{\int_0^\infty \left[ \int_0^{\varepsilon \exp(-x_1^2/2)} dx_2 \right] dx_1} = 2\Phi(C) - 1. \quad (15)$$

which is constant in $\varepsilon$ and can be made arbitrarily close to 1 by choosing a large $C$. So the sequence $\{ P_\varepsilon \}$ is tight.

Next, let $a(x) = (1 + x_1)^{-2}(1 + x_2)^{-2}$ and $g(x) = x$ (see right panel of Figure 2). $a(x)$ is the joint density of two independent Pareto(1) variates. As $x \to \infty$, the tail probability of the Pareto distribution vanishes linearly while $g(x)$ increases linearly. For all $\varepsilon > 0$,

$$\int_{M_\varepsilon} a(x) dx = \int_0^\infty \left[ \int_0^{\varepsilon x_1} (1 + x_2)^{-2} dx_2 \right] (1 + x_1)^{-2} dx_1 = \frac{-\varepsilon(1 - \varepsilon + \log \varepsilon)}{(1 - \varepsilon)^2}. \quad (16)$$
Consider the compact set \( K = [0, C]^2 \). Then for all \( \varepsilon < 1 \),
\[
\int_{M_\varepsilon \cap K^c} a(x)dx = \int_C^\infty \left[ \int_0^{\varepsilon x_1} (1 + x_2)^{-2}dx_2 \right] (1 + x_1)^{-2}dx_1 \\
= -\varepsilon \left[ 1 - \varepsilon - (1 + C) \log \left( \frac{1 + C \varepsilon}{\varepsilon + C \varepsilon} \right) \right] \\
= \frac{1 - \varepsilon - (1 + C) \log \left( \frac{1 + C \varepsilon}{\varepsilon + C \varepsilon} \right)}{(1 + C)(1 - \varepsilon)^2}.
\] (17)

The ratio of (17) and (16) gives the probability of \( P_\varepsilon \{ K^c \} \): As \( \varepsilon \downarrow 0 \),
\[
\frac{\int_{M_\varepsilon \cap K^c} a(x)dx}{\int_{M_\varepsilon} a(x)dx} = \frac{1 - \varepsilon - (1 + C) \log \left( \frac{1 + C \varepsilon}{\varepsilon + C \varepsilon} \right)}{(1 + C)(1 - \varepsilon + \log \varepsilon)} \to 1.
\] (18)

As such, the truncated sequence \( \{ P_\varepsilon \} \) eventually places all the mass outside \( K \) for all \( C \) and thus cannot be tight.

### 3.2 Data Generating Manifold

Given a general DGE \( G : \Upsilon \times \Theta \to Y \), observed data \( y \in Y \), and an \( \varepsilon \geq 0 \), let
\[
G_\varepsilon(y) = \{(u^\top, \theta^\top)^\top \in \Upsilon \times \Theta : \|G(u, \theta) - y\| \leq \varepsilon\}.
\] (19)

For notational succinctness, we write \( G(y) = G_0(y) \). In general, \( G(y) \) may or may not have a positive Lebesgue measure on \( \mathcal{R}^{m+q} \). A further special case of the latter is of interest to us—when \( G(y) \) is a submanifold of \( \Upsilon \times \Theta \subseteq \mathcal{R}^{m+q} \). In this case, we call \( G(y) \) and \( G_\varepsilon(y) \) a data generating manifold and its \( \varepsilon \)-enlargement, respectively. Also let
\[
U_\varepsilon(y) = \{ u \in \Upsilon : (u^\top, \theta^\top)^\top \in G_\varepsilon(y) \text{ for some } \theta \in \Theta \}
\] (20)
be the \( u \)-coordinate projection of \( G_\varepsilon(y) \), and \( U(y) = U_0(y) \). \( G_\varepsilon(y) \) and \( U_\varepsilon(y) \) are the regions of truncation in simulation-based Bayesian and fiducial inference, respectively (see Sections 2.2 and 2.3).

In order to apply Theorem 1 to our inferential context, the following assumptions are made throughout the rest of the paper.

**Assumption 1.** For all \( u \) and \( \theta \) in some neighborhood of \( G(y) \),
(i) \( G : \mathcal{Y} \times \Theta \to \mathbb{R}^n \) is three-time continuously differentiable;

(ii) the \( n \times m \) Jacobian matrix \( \nabla_u G(u, \theta) \) has full row rank, and the \( n \times q \) Jacobian matrix \( \nabla_\theta G(u, \theta) \) has full column rank;

(iii) for a given \( u, \hat{\theta}(y, u) \) defined by (4) is unique.

An immediate consequence of Assumption 1 is the isomorphism of the data generating manifold \( \mathcal{G}(y) \) and its \( u \)-coordinate projection \( U(y) \), which is summarized as Lemma 2 for ease of reference. The proof can be found in Appendix B.

**Lemma 2.** Under Assumption 1, \( \mathcal{G}(y) \subset \mathcal{Y} \times \Theta \subseteq \mathbb{R}^{m+q} \) is isomorphic to \( U(y) \subset \mathcal{Y} \subseteq \mathbb{R}^m \), both of which are twice continuously differentiable submanifolds of dimension \( m + q - n \). In particular, \( U(y) \) can be directly defined as the level set

\[
U(y) = \{ u \in \mathcal{Y} : \nabla_\theta G(u, \hat{\theta}(y, u))^\top (G(u, \hat{\theta}(y, u)) - y) = 0 \},
\]

in which \( \nabla_\theta G(u, \theta) \) is an \( n \times (n - q) \) orthogonal complement of \( \nabla_\theta G(u, \theta) \) that has orthonormal columns and varies smoothly along \( u \) and \( \theta \), and

\[
\mathcal{G}(y) = \{ (u^\top, \hat{\theta}(y, u)^\top)^\top : u \in U(y) \}.
\]

In addition, the intrinsic measures of \( \mathcal{G}(y) \) and \( U(y) \) satisfy

\[
\lambda_{U(y)}(du) = D(u, \theta)^{-1/2} \lambda_{\mathcal{G}(y)}(du, d\theta),
\]

in which

\[
D(u, \theta) = \det \left( \tau_q + \left[ \nabla_\theta G(u, \theta)^\top (\nabla_u G(u, \theta) \nabla_u G(u, \theta)^\top) \right]^{-1} \nabla_\theta G(u, \theta) \right)^{-1}.
\]

### 3.3 Bayesian and Fiducial Inference

Before presenting our geometric characterization, we establish the following representation of the likelihood function.
Lemma 3. The likelihood function \( f(y|\theta) \) can be expressed by

\[
f(y|\theta) = \int \frac{\rho(u)}{\det (\nabla_u G(u, \theta) \nabla_u G(u, \theta)^	op)^{1/2}} d\lambda_{G_\theta(y)}, \quad (25)
\]

in which \( G_\theta(y) = \{ u \in \mathcal{Y} : G(u, \theta) = y \} \), termed the \( \theta \)-section of \( G(y) \), is a \((m - n)\)-dimensional submanifold of \( \mathcal{Y} \).

Proof. For a fixed \( \theta \), \( G(\cdot, \theta) \) is a submersion by (ii) of Assumption 1; therefore, the level set \( G_\theta(y) \) is a submanifold of \( \mathcal{Y} \) of dimension \( m - n \) \citep[Corollary 5.13]{Lee2013}. (25) follows from the smooth coarea formula \citep[e.g., Section III.8]{Chavel2006}:

\[
P\{ Y \in B | \theta \} = P\{ G(U, \theta) \in B | \theta \} = \int_{G(U, \theta) \in B} \rho(u) du = \int_B \left\{ \int \frac{\rho(u)}{\det (\nabla_u G(u, \theta) \nabla_u G(u, \theta)^	op)^{1/2}} d\lambda_{G_\theta(y)} \right\} dy \quad (26)
\]

for any measurable \( B \subseteq \mathcal{Y} \).

Remark 4. So far, we have seen three different ways to interpret the dimension of the data generating manifold, i.e., \( m + q - n \).

1) \((m + q) - n\): Most obviously, the data generating manifold \( G(y) \) is a submanifold of the \((m + q)\)-dimensional space \( \mathcal{Y} \times \Theta \) that is implicitly defined by the \( n \)-dimensional constraint \( G(u, \theta) = y \).

2) \(m - (n - q)\): By Lemma 2, \( G(y) \) is isomorphic to its \( u \)-coordinate projection \( U(y) \), which is a submanifold of the \( m \)-dimensional space \( \mathcal{Y} \) that is implicitly defined by the \((n - q)\)-dimensional constraint \( \nabla_{\theta} G(u, \hat{\theta}(y, u))^	op (G(u, \hat{\theta}(y, u)) - y) \).

3) \((m - n) + q\): By Lemma 3, the \( \theta \)-section \( G_\theta(y) \) is a \((m - n)\)-dimensional submanifold of \( \mathcal{Y} \), and \( G(y) = G_\theta(y) \times \Theta \).

ABC typically operates on the joint space \( \mathcal{Y} \times \Theta \) and thus naturally adopts the first view. Meanwhile, the second view is aligned with Hannig et al.'s \citeyear{Hannig2016} treatment of GFI on the space \( \mathcal{Y} \). As will be elaborated in the next two propositions, the third view links our geometric perspective back to the conventional definitions of Bayesian posteriors and
We are now ready to represent Bayesian posteriors and GFDs as limits on either \( G(y) \) or \( U(y) \). We consider Bayesian inference first in the next proposition; the proof can be found in Appendix C.

**Proposition 4.** For Bayesian inference, the general notations of Theorem 1 reduce to \( x = (u^\top, \theta^\top)^\top \), \( X = \mathcal{Y} \times \Theta \), \( \mathcal{M}_\varepsilon = G_\varepsilon(y) \), \( \mathcal{M} = G(y) \), \( a(x) = \pi(\theta)\rho(u) \), and \( h(x) = G(u, \theta) - y \). Under the assumptions of Theorem 1, the weak limit of (2) as \( \varepsilon \downarrow 0 \) has the following absolutely continuous density

\[
f_B(u, \theta) \propto \rho(u)\pi(\theta) \det \left( \nabla_u G(u, \theta) \nabla_G u(u, \theta)^\top + \nabla_\theta G(u, \theta) \nabla_G \theta(u, \theta)^\top \right)^{-1/2}
\]

with respect to \( \lambda_{G(y)} \). Equivalently, the limit can be characterized by the density

\[
\tilde{f}_B(u) \propto \rho(u)\pi(\theta) \det \left( \nabla_u G(u, \theta) \nabla_G u(u, \theta)^\top \right)^{-1/2} \\
\cdot \det \left( \nabla_\theta G(u, \theta)^\top \left( \nabla_u G(u, \theta) \nabla_G u(u, \theta)^\top \right)^{-1} \nabla_\theta G(u, \theta) \right)^{-1/2}
\]

with respect to \( \lambda_{U(y)} \). Moreover, the density of \( \hat{\theta}(y,u) \) under (28), or equivalently the \( \theta \)-marginal of (27), is proportional to \( \pi(\theta)f(y|\theta) \).

Proposition 5 gives a similar characterization for fiducial distributions; the proof can also be found in Appendix C.

**Proposition 5.** For GFI, the general notations of Theorem 1 reduce to \( x = u \), \( X = \mathcal{Y} \), \( \mathcal{M}_\varepsilon = U_\varepsilon(y) \), \( \mathcal{M} = U(y) \), \( a(x) = \rho(u) \), and \( h(x) = \nabla_\theta G(u, \hat{\theta}(y,u))^\top (G(u, \hat{\theta}(y,u)) - y) \).

Under the assumptions of Theorem 1, the weak limit of (3) as \( \varepsilon \downarrow 0 \) has the following absolutely continuous density

\[
\tilde{f}_F(u) \propto \rho(u) \det \left( \nabla_\theta G(u, \hat{\theta}(y,u))^\top \nabla_u G(u, \hat{\theta}(y,u)) \nabla_u G(u, \hat{\theta}(y,u))^\top \nabla_\theta G(u, \hat{\theta}(y,u)) \right)^{-1/2}
\]

(29)
with respect to $\lambda_{U(y)}$. Equivalently, the limit can be characterized by the density

$$f_F(u, \theta) \propto \rho(u) \det \left( \nabla_\theta G(u, \theta)^\top \nabla_\theta G(u, \theta) \right)^{1/2} 
\cdot \det \left( \nabla_u G(u, \theta) \nabla_u G(u, \theta)^\top + \nabla_\theta G(u, \theta) \nabla_\theta G(u, \theta)^\top \right)^{-1/2}$$

(30)

with respect to $\lambda_{G(y)}$. Moreover, the density of $\hat{\theta}(y, u)$ under (30), or equivalently the $\theta$-marginal of (29), is proportional to

$$\int \rho(u) \cdot \det \left( \nabla_\theta G(u, \theta)^\top \nabla_\theta G(u, \theta) \right)^{1/2} \det \left( \nabla_u G(u, \theta) \nabla_u G(u, \theta)^\top \right)^{-1/2} d\lambda_{G(y)}(u).$$

(31)

**Remark 5.** Under Assumptions A.1–A.4 of Hannig et al. (2016), the $\theta$-section of the data generating manifold, i.e., $G_\theta(y)$, reduces to the single point $\hat{u}(y, \theta)$ (see Section 2.3). The likelihood representation (25) then becomes

$$f(y|\theta) = \rho(\hat{u}(y, \theta)) \cdot \det \left( \nabla_u G(\hat{u}(y, \theta), \theta) \nabla_u G(\hat{u}(y, \theta), \theta)^\top \right)^{-1/2},$$

which appeared on p. 2 in the supplementary document of Hannig et al. (2016). Hence, (31) is simplified to (7).

**Remark 6.** By the Matrix Determinant Lemma, the second determinant on the right-hand side of (30) can be alternatively expressed as

$$\det \left( \nabla_u G(u, \theta) \nabla_u G(u, \theta)^\top \right)^{1/2} 
\cdot \det \left( \iota_q + \nabla_\theta G(u, \theta)^\top \left[ \nabla_u G(u, \theta) \nabla_u G(u, \theta)^\top \right]^{-1} \nabla_\theta G(u, \theta) \right)^{1/2},$$

(32)

in which $\iota_q$ is a $q \times q$ identity matrix. (32) is often computationally more efficient for the reasons that $\nabla_u G(u, \theta) \nabla_u G(u, \theta)^\top$ can be highly sparse and structured (e.g., block diagonal in the repeated-measures ANOVA example; see Section 5), and that the second determinant is computed with a small $q \times q$ matrix.

**Remark 7.** In the light of Theorem 1, (30) can be thought as restricting the ambient
density

\[ a_F(u, \theta) = \rho(u) \det \left( \nabla_\theta G(u, \theta)^\top \nabla_\theta G(u, \theta) \right)^{1/2} \]

to the \( y \)-generating manifold, while Bayesian inference concerns restricting

\[ a_B(u, \theta) = \rho(u)\pi(\theta) \]

to the same manifold. On the one hand, GFI and Bayesian inference share the same feature that only the data generating manifold but not the ambient density depends on the observed data \( y \). In both frameworks, “model specification” boils down to defining an ambient density on the joint space \( \Upsilon \times \Theta \), “model fitting” is translated into restricting the ambient density to the \( y \)-generating manifold, and “inference about \( \theta \)” and “prediction about \( u \)” are generated from the marginals of the limiting density. On the other hand, the determinant term in \( a_F \) depends on both \( u \) and \( \theta \), while the prior \( \pi(\theta) \) in \( a_B \) is a function of \( \theta \) only. As such, GFI does not reduce to Bayesian inference in general. The discrepancy is rooted in whether \( \theta \) can be freely generated based on “prior beliefs” (Bayesian) or has to be determined by solving an optimization problem that varies along \( u \) (fiducial).

4 Review of Markov Chain Monte Carlo Sampling on Manifolds

Monte Carlo approximations to a fiducial or a Bayesian posterior distribution—when viewed as an absolutely continuous distribution defined on a smooth manifold—can be constructed via manifold MCMC sampling. In this section, we review a manifold random-walk Metropolis (RWM) algorithm proposed by Zappa et al. (2018). We focus on a specific Gaussian proposal that corresponds to a one-step discretization of the constrained overdamped Langevin process (Lelièvre et al., 2012, Section 3.3). For generality, we adopt the notation of Theorem 1 in the current section. The algorithms are presented assuming that \( M \) is unbounded, though incorporating additional inequality constraints is straightforward (see, e.g., Remark 6 of Lelièvre et al., 2019).

\footnote{Data-dependent priors, which are sometimes allowed in empirical and objective Bayes, are not part of this discussion.}
Algorithm 1 Manifold Random-Walk Metropolis Update

**input:** Initial value \( x \in \mathcal{M} \), target density \( f : \mathcal{R}^d \to \mathcal{R} \), proposal parameters \( \mu(x) \in \mathcal{R}^{d-n} \), \( \Sigma(x) \in \mathcal{R}^{(d-n)\times(d-n)}_+ \), and \( \delta > 0 \), tuning parameters \( \gamma \) and \( R \) for Project (Algorithm 2)

1: Sample \( z \) from \( \mathcal{N}(\mu(x), \delta \Sigma(x)) \) and set \( w = \nabla h(x)z \)
2: Propose \( x' = \text{Project}(x + w, \nabla h(x), \gamma, R) \)
3: if fail to find \( x' \) then
4: return \( x \)
5: end if
6: Compute \( z' = \nabla h(x')^\top (x - x') \) and \( w' = \nabla h(x')z' \)
7: Set \( x'' = \text{Project}(x' + w', \nabla h(x'), \gamma, R) \)
8: if fail to find \( x'' \) or \( x'' \neq x \) then
9: return \( x \)
10: end if
11: Compute \[ \alpha(x; x') = \min \left\{ 1, \frac{f(x') \phi(z'; \mu(x'), \delta \Sigma(x'))}{f(x) \phi(z; \mu(x), \delta \Sigma(x))} \right\} \] (33)
12: Sample \( u \in [0, 1] \) from Unif[0, 1]
13: if \( u \leq \alpha(x; x') \) then
14: return \( x' \)
15: else
16: return \( x \)
17: end if

Algorithm 2 Project: Projection to Manifold \( \mathcal{M} \) along \( B \)

**input:** Initial location \( x_0 \in \mathcal{R}^d \), full-rank basis matrix \( B \in \mathcal{R}^{d\times n}_+ \), convergence tolerance \( \gamma > 0 \), maximum number of iterations \( R \)

1: Set \( a_0 = 0_n \), where \( 0_n \) is a \( n \times 1 \) vector of zeros
2: for \( r = 0, \ldots, R - 1 \) do
3: if \( \|h(x_r)\| \leq \gamma \) then
4: return \( x_r \)
5: end if
6: Update \( a_{r+1} = a_r - \left[ \nabla h(x_r + Ba_r)^\top B \right]^{-1} h(x + Ba_r) \)
7: Compute \( x_{r+1} = x_0 + Ba_r \)
8: end for
9: Throw an error

4.1 Manifold Random-Walk Metropolis

The pseudocode for a single manifold RWM update is summarized in Algorithm 1. With a slight abuse of notation, \( f \) in the pseudocode denotes a smooth extension of the target density \( f \) (with respect to \( \lambda_\mathcal{M} \)) to the ambient space \( \mathcal{R}^d \). Given an initial value \( x \) on the manifold \( \mathcal{M} \), a proposal \( x' \) is generated from a random walk on the tangent space at \( x \)
(Line 1), followed by a projection back to the manifold along the normal direction\(^7\) (Line 2). Let \( T_x \mathcal{M} = \{ w \in \mathbb{R}^d : \nabla h(x) w = 0 \} \) be the tangent space of \( \mathcal{M} \) at \( x \), and \( \nabla h(x) \in \mathbb{R}^{d \times (d-n)} \) be an orthogonal complement of \( \nabla h(x)^\top \in \mathbb{R}^{d \times n} \) with orthonormal columns. \( \nabla h \) forms a basis for \( T_x \mathcal{M} \). The random-walk step entails generating \( w = \nabla h(x) z \in T_x \mathcal{M} \), in which \( z \) follows \( \mathcal{N}(\mu(x), \delta \Sigma(x)) \), \( \mu(x) \in \mathbb{R}^{d-n} \), \( \Sigma(x) \in \mathbb{R}^{(d-n) \times (d-n)} \) is positive definite, and \( \delta > 0 \) is the proposal scale parameter. The point \( x + w \) resulted from the random walk needs to be retracted back to \( \mathcal{M} \) to yield a valid proposal. In particular, we find a coefficient vector \( a \in \mathbb{R}^n \) that solves \( h(x + z + \nabla h(x)a) = 0 \). Because the constraint function \( h \) is generally nonlinear, we follow Zappa et al. (2018) to apply a standard Newton solver. This projection step is abbreviated as \textbf{Project} in the pseudocode: The four arguments required by the function call of \textbf{Project} are described in the input line of Algorithm 2.

The proposal \( x' \) is not accepted unless it passes all the following three checks. First, it is possible that the function \textbf{Project} throws an error—or equivalently, the Newton solver fails to converge (see Lines 3\textendash 5 of Algorithm 1); if so, we have to revert to the original \( x \) and proceed to the next cycle. Second, we need to confirm that a reverse move starting from \( x' \) recovers the original point \( x \) (Lines 6\textendash 10); a graphical illustration for the potential failure of such a reversal move can be found in Figure 2 of Lelièvre et al. (2019). Finally, a standard Metropolis-Hastings step is performed (Lines 11\textendash 17), in which the acceptance ratio is given by (33). It was shown in Zappa et al. (2018) that the above RWM update satisfies the detailed balance condition when the equations were solved exactly in the retraction steps (Lines 2 and 7). When a numerical solver is employed, which is typically the case in practice, the manifold RWM algorithm can be understood as a noisy MCMC method (Alquier et al., 2016).

\(^7\) We follow Zappa et al. (2018) to call the operation a “projection”; however, it is different from an orthogonal projection to the manifold.
4.2 Proposal Distribution

We found in pilot experiments that a Gaussian proposal (Line 1) with
\[
\mu(x)^\top = \frac{\delta^2}{2} \nabla \log f(x) \nabla h(x)
\]
and \(\Sigma(x) \equiv \iota_{(d-n) \times (d-n)}\) fares efficient even when the dimension of the manifold (i.e., \(d - n\)) is high. The corresponding manifold RWM update yields an Euler discretization of the constrained overdamped Langevin diffusion (with an identity mass matrix; Lelièvre et al., 2012, Proposition 3.6), which is also equivalent to a single update of the “location variable” while simulating the constrained Hamilton dynamics via the RATTLE discretization \(^8\) (Brubaker et al., 2012; Lelièvre et al., 2019). In case the gradient of the log density is challenging to evaluate (e.g., the gradient of the log fiducial density (30) involves the second derivatives of the DGE), we may substitute \(\nabla \log f(x)\) in (34) by a numerical estimate. In fact,
\[
\nabla \log f(x) \nabla h(x) = \nabla_t \log f(x + \nabla h(x)^\top t)|_{t=0_{d-n}}. \quad (35)
\]
Compared to differentiating \(\log f\) with respect to \(x \in \mathcal{R}^d\), the right-hand side derivative in (35) is taken with respect to the lower-dimensional \(t \in \mathcal{R}^{d-n}\) and thus can be more economical to numerically approximate.

5 Example: Repeated-Measures ANOVA

Next, we apply our main results to a repeated-measures ANOVA example, which is a mixed-effects model. GFI for Gaussian linear mixed-effects models has been studied by Cisewski and Hannig (2012); however, their development is confined to the “fat data” setting (i.e., (5) with a positive tolerance \(\varepsilon\)), and the proposed sequential Monte Carlo algorithm suffers from numerical degeneracy when \(\varepsilon\) is small. From the new geometric perspective, we can not only express the exact fiducial density (i.e., the weak limit as \(\varepsilon \downarrow 0\))

\(^8\) One can extend Algorithm [1] to a manifold Hamiltonian Monte Carlo sampler by repeatedly executing Lines [1-10] with a small, fixed “time increment” \(\delta\).
in closed-form but also generate fiducial samples conveniently using manifold MCMC algorithms.

5.1 Model

In a within-subject design, let \(X_{ij}\) denote the observed response of subject \(j\) in condition \(i\), where \(i = 1, \ldots, I\) and \(j = 1, \ldots, J\) with \(I, J > 1\). Repeated-measures ANOVA decomposes each response entry \(X_{ij}\) into the sum of the treatment mean \(\mu_i\), subject effect \(\sigma_z Z_j\), and the interaction effect \(\sigma_e E_{ij}\):

\[
X_{ij} = \mu_i + \sigma_z Z_j + \sigma_e E_{ij},
\]

(36)
in which \(Z_j\) and \(E_{ij}\) are continuous random variables, and \(\sigma_z\) and \(\sigma_e\) are the respective scale parameters. (36) amounts to the component-wise expression of the DGE. To be consistent with our generic notation, identify

\[
Y = \text{vec}(X), \quad U = (Z^\top, \text{vec}(E)^\top)^\top, \quad \theta = (\mu^\top, \sigma_z, \sigma_e)^\top,
\]

(37)
in which \(X = \{X_{ij}\}\), \(Z = \{Z_j\}\), \(E = \{E_{ij}\}\), and \(\mu = \{\mu_i\}\). The dimensions of \(Y\), \(U\), and \(\theta\) are \(n = IJ\), \(m = IJ + J\), and \(q = I + 2\), respectively.

In the next proposition, we verify the crucial tightness assumption that allows us to apply Proposition [3]. The proof can be found in Appendix [D].

**Proposition 6.** Under a repeated-measures ANOVA model, suppose that \(\text{vec}(E)\) and \(Z\) are independent, spherically distributed random vectors on \(\mathcal{R}^{IJ}\) and \(\mathcal{R}^J\), respectively. Then the collection of probability measures (35) indexed by \(\varepsilon\) is tight.

**Remark 8.** The additional distributional assumption for \(\text{vec}(E)\) and \(Z\) is made for ease of theoretical justification. A spherically distributed variate \(S\) is subject to a unique factorization \(S = RV\), where \(V \in \mathcal{R}^{IJ}\) is uniform on the unit sphere and \(R > 0\) is a positive, continuous random variable ([Fang et al., 1990]). Common examples of spherical distributions are multivariate Gaussian and \(t\) distributions, which are popular choices of error distributions for linear models ([Fraser and Ng, 1980]).
Remark 9. The proof of Proposition 6 in Appendix D can be adapted to handle unbalanced designs (i.e., $i = 1, \ldots, I$ where $I_j$’s may not be identical for different $j$’s) or even more general linear mixed-effects models considered by Cisewski and Hannig (2012). We only need to modify in (69) the definition of “the coefficient vector” $\beta$ to include all fixed effects and scale parameters for random effects, and correspondingly the definition of the “design matrix” $W(z)$.

Pointwise evaluation of the fiducial density (30) requires formulas for the Jacobian matrices. Under the repeated-measures ANOVA model, $\nabla_u G(u, \theta)$ and $\nabla_\theta G(u, \theta)$ have blocked matrix representations corresponding to the partitions of $U$ and $\theta$ in (37):

$$\nabla_u G(u, \theta) = (\iota_J \otimes \sigma_z 1_I : \sigma_e \iota_{IJ}),$$
$$\nabla_\theta G(u, \theta) = (1_J \otimes \iota_I : Z \otimes 1_I : \text{vec}(E)).$$

(38)

Note that the dimensions of $\nabla_u G$ and $\nabla_\theta G$ are $IJ \times (IJ + J)$ and $IJ \times (I + 2)$. Naively evaluating the fiducial density and applying the manifold MCMC update incur matrix operations up to $O(I^3J^3)$ complexity, which can be prohibitively expensive when $I$ or $J$ is large. In Appendix E, we demonstrate that the computational cost can be reduced to $O(I^3J)$ thanks to the specific structure of (38).

5.2 Empirical Data: Orthodontic Growth

Using the orthodontic growth data (Potthoff and Roy, 1964), we apply the manifold RWM algorithm to sample from the fiducial distribution with density (30). The data set contains measures of the distance between the pituitary and pterygomaxillary fissures for a total number of 27 children, including 16 males and 11 females. Measures were obtained every two years from age 8 to 14, resulting in four measures per child. We built models for the female and male subjects separately, implicating that all the parameters are potentially different across gender groups. The distributions of random components are set to $Z \sim \mathcal{N}(0_I, \iota_I)$ and $\text{vec}(E) \sim t_{10}(0_{IJ}, \iota_{IJ})$.

The step size of the Gaussian random walk was tuned such that the empirical
Figure 3: Trace plots (top panels) and autocorrelation plots (bottom panels) for the manifold random-walk Metropolis sampler in the orthodontic growth example. Results are shown separately for the male and female sub-samples (in different colors) and for each of the six parameters (in columns). We retained 10000 cycles after burning in the first 10000 to remove the impact of arbitrary starting values. On the trace plots, the horizontal solid lines indicate the means of the Monte Carlo samples.

The acceptance rate is about 0.5 in pilot runs. The proposal scale parameter $\delta$ was set to 1.03 and 1.10 in the male and female sub-samples, which results in empirical acceptance rates 0.4906 and 0.5082 out of 10000 retained MCMC cycles in the end. The tolerance and the maximum iterations of the Newton solver were set to $10^{-6}$ and 50, respectively. The manifold RWM sampler was implemented in MATLAB (2021) and is available upon request.

Trace and autocorrelation plots of the generated Markov chains were displayed for each of the six parameters in Figure 3. Note that we plot the logarithms of the scale parameters $\sigma_z$ and $\sigma_e$. The plots were generated based on 10000 retained cycles after discarding the first 10000 cycles to remove the influence of arbitrary starting states. All the twelve reported univariate sample paths appear to be stationary, and the effective sample sizes (Gelman et al., 2013, Chapter 11) range from 314.5156 to 1896.7478.

It is concluded that measures for males are on average larger than those for females across all age levels. However, there appears no substantial between-gender discrepancy in
individual variability: The gender-specific fiducial distributions of the scale parameter $\sigma_z$ are largely overlapping. Meanwhile, interaction effects or residual terms for males are more variable compared to those for females (in terms of the scale parameter $\sigma_e$).

6 Concluding Remarks

In the present paper, we approach GFI, as well as Bayesian inference for comparison, from a differential geometric perspective. Conditional on the observed data, a statistical model with a smooth DGE (meeting the assumptions detailed in Section 3) defines a manifold (i.e., the data generating manifold) embedded in the joint space of random components and parameters. We show that a Bayesian posterior and the GFD defined by Hannig et al. (2016) can both be represented as closed-form densities with respect to the intrinsic measure of the data generating manifold. In the meantime, GFD is generally not equivalent to any Bayesian posterior, which is evident from a factorization of their density function (see Remark 7). We also demonstrate that manifold MCMC samplers can be utilized to construct Monte Carlo approximations to the GFD in an empirical example.

Taking an alternative, yet still differential geometric, perspective on GFI, Murph et al. (2022b) defined a GFD whose parameter space $\Theta$ itself is a manifold. In contrast to this paper, where the GFD is defined as the limiting measure of an ambient distribution constrained to a sequence of shrinking sublevel sets (an extrinsic perspective), Murph et al. (2022b) define their distribution directly on the manifold (an intrinsic perspective) using the smooth local structure. Whenever the dimension of the random component and the observed data are the same ($m = n$), Theorem 3.1 from Hwang (1980) can be used to calculate an extrinsic analogue of the GFD from Murph et al. (2022b). Under some regularity conditions, Murph et al. (2022b) showed that these extrinsic and intrinsic perspectives converge in the local limit. A natural extension of the main result of this paper is to extend Proposition 5 to additionally handle a constrained parameter space, which can be seen as both a generalization of the result from Murph et al. (2022b) (for $m > n$), and as an alternative, extrinsic perspective using limiting measures.
Appendix A

Proof of Theorem 1

The proof proceeds in two stages. In Section A.1 we first establish the claim of Theorem 1 under the additional assumption that $\mathcal{M}_\varepsilon$ is compact for all sufficiently small $\varepsilon$. The result is then extended to non-compact cases (Section A.2).

A.1 Compact Case

Suppose that there exists $\varepsilon_0 > 0$ such that $\mathcal{M}_\varepsilon$ is compact whenever $\varepsilon \leq \varepsilon_0$—this also implies that the manifold $\mathcal{M}$ is compact. Also let $N_x\mathcal{M}$ be the normal space of $\mathcal{M}$ at $x \in \mathcal{M}$, and $\mathcal{N}\mathcal{M} = \{(x, v) : x \in \mathcal{M}, v \in N_x\mathcal{M}\}$ be the normal bundle of $\mathcal{M}$. By the Tubular Neighborhood Theorem (Lee, 2013, Corollary 6.17) and the compactness of $\mathcal{M}$, there exists a $\tau > 0$ and a tubular neighborhood of $\mathcal{M}$ that is defined as the diffeomorphic image of the open set $V = \{(x, v) \in \mathcal{N}\mathcal{M} : \|v\| < \tau\}$ under the map $\chi : \mathcal{R}^d \times \mathcal{R}^d \to \mathcal{R}^d$, $\chi(x, v) \mapsto x + v$. We denote such a tubular neighborhood by $T = \chi(V)$. Our goal is to show that, for all bounded continuous function $g : \mathcal{R}^d \to \mathcal{R}$,

$$
\varepsilon^{-n} \int_{\mathcal{M}_\varepsilon} g(x)a(x)dx \to \frac{\pi^{n/2}}{\Gamma(n+1/2)} \int g(x)a(x) \det \left( \nabla h(x)\nabla h(x)^T \right)^{-1/2} \lambda_\mathcal{M}(dx)
$$

(39)

as $\varepsilon \downarrow 0$. Because the constant function 1 is also bounded and continuous, the desired result (12) follows from (39) and the Portmanteau Lemma.

We first note that $\mathcal{M}_\varepsilon$ is eventually contained in the closure of $T$, denoted $\overline{T}$, as $\varepsilon$ decreases to 0. To see this, take an $x \in \mathcal{M}_\varepsilon$, $\varepsilon \leq \varepsilon_0$, and let $\zeta(x) = \arg \min_{x' \in \mathcal{M}} \|x - x'\|$. By the Mean Value Theorem,

$$
\|h(x)\| = \|h(x) - h(\zeta(x))\| = \|\nabla h(\bar{x})(x - \zeta(x))\|,
$$

(40)

where $\bar{x}$ is between $x$ and $\zeta(x)$. Because $h$ is a submersion, $\nabla h(\bar{x})$ has full column rank, and hence the smallest singular value of $\nabla h(\bar{x})$, denoted $\sigma_n(\bar{x})$, is positive. (40) implies that $\|x - \zeta(x)\| \leq \|h(x)\|/\sigma_n(\bar{x}) \leq \varepsilon/\sigma_n(\bar{x})$. By the compactness of $\mathcal{M}_\varepsilon$ and $\mathcal{M}$, we have $\sigma_\varepsilon = \inf_{x \in \mathcal{M}_\varepsilon} \sigma_n(\bar{x}) > 0$ and consequently $\|x - \zeta(x)\| \leq \varepsilon/\sigma_\varepsilon$. Because $\sigma_\varepsilon$ is non-decreasing
as \( \varepsilon \downarrow 0, \varepsilon/\sigma_\varepsilon \) is eventually smaller than \( \tau \). Without loss of generality, suppose henceforth that \( \mathcal{M}_\varepsilon \) is contained in \( \mathcal{T} \).

By definition, any \( x \in \mathcal{T} \) is subject to a unique decomposition \( x = s + Q(s)t \) where \( s \in \mathcal{M}, t \) in some compact neighborhood of 0, and \( Q(s) \) is a \( d \times n \) orthonormal basis matrix for the range of \( \nabla h(s)^\top \) that varies smoothly in \( s \). Introduce the shorthand notation \( a(s,t) = a(s+Q(s)t) \); \( g(s,t) \) and \( h(s,t) \) are similarly defined. Following [Weyl (1939)], we have

\[
\int_{\mathcal{M}_\varepsilon} g(x) a(x) dx = \int \left[ \int_{\mathcal{M}_\varepsilon} g(s,t) a(s,t) J(s,t) dt \right] \lambda_\mathcal{M} (ds). \tag{41}
\]

In (41), \( \mathcal{M}_\varepsilon^s = \{ t \in \mathcal{R}^n : \|h(s,t)\| \leq \varepsilon \} \) is the \( s \)-section of \( \mathcal{M}_\varepsilon \); \( J(s,t) = \det(K(s,t)) \), and the \( (d-n) \times (d-n) \) matrix \( K(s,t) \) has elements \( K_{\alpha\beta}(s,t) = \mathbb{I}_{\{\alpha=\beta\}} + \sum_{i=1}^n t_i G^\beta_\alpha(i; s) \), where \( \alpha, \beta \in \{1, \ldots, d-n\} \), \( t_i \) is the \( i \)-th coordinate of \( t \), and \( G^\beta_\alpha(i; s) \) denotes the coefficients of the second fundamental form in the \( i \)-th direction of the normal space at \( s \).

While the detailed expression of \( J(s,t) \) is not used in the sequel, we do need the obvious fact that \( J(s,0) = \det(t_{d-n}) = 1 \).

Now fix \( s \in \mathcal{M} \). By the twice continuous differentiability of \( h \), we have the Taylor series expansion of \( h(s,t) \) at \( t = 0 \):

\[
h(s,t) = H_0(s)t + \xi(s,t), \tag{42}
\]

in which

\[
H_t(s) = \nabla_t h(s,t) = \nabla h(s + Q(s)t)Q(s) \tag{43}
\]

is the directional derivative of \( h \) along \( t \), and the remainder \( \xi(s,t) \) satisfies \( \|\xi(s,t)\| = o(\|t\|) \) (uniformly in \( s \)). Let \( \hat{\mathcal{M}}_\varepsilon^s = \{ t \in \mathcal{R}^n : \|H_0(s)t\| \leq \varepsilon \} \), which is an ellipsoid in \( \mathcal{R}^n \). Because \( H_0(s) \) is of full rank and \( \mathcal{M} \) is compact, \( \|H_0(s)t\| \) is bounded from below by a constant multiple of \( \|t\| \); therefore, \( \hat{\mathcal{M}}_\varepsilon^s \) is contained in the tubular
neighborhood when \( \varepsilon \) is sufficiently small. By the volume formula for ellipsoids,

\[
\text{vol}\{\mathcal{M}_\varepsilon\} = \int_{\mathcal{M}_\varepsilon} dt = \frac{\pi^{n/2}\varepsilon^n}{\Gamma(1 + n/2)} \det (H_0(s)^\top H_0(s))^{-1/2} \\
= \frac{\pi^{n/2}\varepsilon^n}{\Gamma(1 + n/2)} \det (\nabla h(s)\nabla h(s)^\top)^{-1/2}.
\]

The last equality of (44) follows from

\[
\det (H_0(s)^\top H_0(s)) = \det (Q(s)^\top Q(s) R(s) R(s)^\top Q(s)^\top Q(s)) = \det (R(s) R(s)^\top) \\
= \det (R(s)^\top R(s)) = \det (\nabla h(s)\nabla h(s)^\top),
\]

in which \( \nabla h(s)^\top = Q(s) R(s) \) is the expansion of \( \nabla h(s)^\top \) on the basis \( Q(s) \), and

\( H_0(s) = \nabla h(s) Q(s) \) by (43). Our remaining task is to show that

\[
\varepsilon^{-n} \int_{\mathcal{M}_\varepsilon} g(s,t)a(s,t) J(s,t) dt \to \frac{\pi^{n/2}}{\Gamma(1 + n/2)} g(s,0)a(s,0) \det (\nabla h(s)\nabla h(s)^\top)^{-1/2}
\]

for each \( s \) in the interior of \( \mathcal{M} \) and then argue that (39) is a consequence of (46) and the Dominated Convergence Theorem.

To establish (46), note that

\[
\varepsilon^{-n} \left| \int_{\mathcal{M}_\varepsilon} g(s,t)a(s,t) J(s,t) dt - g(s,0)a(s,0) \int_{\mathcal{M}_\varepsilon} dt \right| \\
\leq \varepsilon^{-n} \int_{\mathcal{M}_\varepsilon} |g(s,t)a(s,t) J(s,t) - g(s,0), a(s,0)| dt + \varepsilon^{-n} \int_{\mathcal{M}_\varepsilon \Delta \mathcal{M}_\varepsilon} |g(s,t)a(s,t) J(s,t)| dt,
\]

in which \( A \Delta B \) stands for the symmetric difference between sets \( A \) and \( B \). For reasons that \( a(s,t), g(s,t), \) and \( J(s,t) \) are all continuous at \( t = 0 \) and that the ellipsoid \( \mathcal{M}_\varepsilon \) is compact,

\[
(1) \leq \sup_{t \in \mathcal{M}_\varepsilon} |g(s,t)a(s,t) J(s,t) dt - g(s,0)a(s,0)| \frac{\pi^{n/2}}{\Gamma(1 + n/2)} \det (\nabla h(s)\nabla h(s)^\top)^{-1/2} \to 0.
\]

(48)
For (II), the Taylor expansion (42) allows us to find \( l(\varepsilon) \leq \varepsilon \leq u(\varepsilon) \) such that
\[
\begin{align*}
    u(\varepsilon) - l(\varepsilon) &= o(\varepsilon), \\
    u(\varepsilon) &\leq \varepsilon_0, \text{ and } \mathcal{M}_{s(\varepsilon)}^s \subseteq \mathcal{M}_s^s \subseteq \hat{\mathcal{M}}_{u(\varepsilon)}^s \text{ for sufficiently small } \varepsilon. \text{ We then have}
\end{align*}
\]
\[
(II) \leq \varepsilon^{-n} \int_{\mathcal{M}_{u(\varepsilon)}^s \setminus \mathcal{M}_{l(\varepsilon)}^s} |g(s, t) a(s, t) J(s, t)| \, dt \leq \sup_{t \in \mathcal{M}_{s_0}^s} |g(s, t) a(s, t) J(s, t)| \cdot \varepsilon^{-n} \int_{\hat{\mathcal{M}}_{u(\varepsilon)}^s \setminus \mathcal{M}_{l(\varepsilon)}^s} \, dt. \hspace{1cm} (49)
\]
On the right-hand side of (49), the supremum is bounded by continuity and compactness; the remaining part satisfies
\[
\varepsilon^{-n} \int_{\mathcal{M}_{u(\varepsilon)}^s \setminus \mathcal{M}_{l(\varepsilon)}^s} \, dt = \frac{(u(\varepsilon)^n - l(\varepsilon)^n) \pi^{n/2}}{\varepsilon^n n!} \lambda^{(n+1/2)} (\nabla h(s) \nabla h(s)^\top)^{-1/2} \rightarrow 0. \hspace{1cm} (50)
\]
As such, (II) → 0 as \( \varepsilon \downarrow 0 \). together with (48), we have now established (46).

The boundary of \( \mathcal{M} \), if exists, is a submanifold of dimension \( d - n - 1 \), which has measure 0 under \( d\mathcal{M} \); furthermore, the right-hand side of (46) remains continuous and bounded at the boundary of \( \mathcal{M} \) by assumption. Applying the Dominated Convergence Theorem concludes the proof of the compact case.

**A.2 Non-Compact Case**

The same limit can be established even if \( \mathcal{M}_\varepsilon \) is not compact for all \( \varepsilon > 0 \). By Prohorov’s Theorem, the tightness of \( \{P_\varepsilon\} \), i.e., Assumption (iii), guarantees that \( \{P_\varepsilon\} \) contains converging subsequences. Let \( \{P_{\varepsilon_k}\} \) be a subsequence such that \( \varepsilon_k \downarrow 0 \) and \( P_{\varepsilon_k} \rightarrow P^* \) as \( k \rightarrow \infty \). We first show that \( 0 < \int a(x) \det (\nabla h(x) \nabla h(x)^\top)^{-1/2} \lambda_{\mathcal{M}}(dx) < \infty \), and consequently the density (12) corresponds to a proper probability measure \( P_0 \). The positivity part follows directly from Assumptions (i) and (ii). To establish the finiteness part, take any bounded \( P^* \)-continuity set \( B \) such that \( P^*\{B\} = P^*\{\bar{B}\} > 0 \). Again by tightness, we can find for each \( \eta \in (0, 1/2) \) a compact set \( K_\eta \subset \mathcal{R}^d \) such that
\[
P_{\varepsilon_k}\{K_\eta\} \geq 1 - \eta \text{ and } \bar{B} \subseteq K_\eta \text{ for all } k. \text{ Then}
\]
\[
\left| \frac{\int_{\bar{B}} a(x) \det (\nabla h(x) \nabla h(x)^\top)^{-1/2} \lambda_{\mathcal{M}}(dx)}{\int_{K_\eta} a(x) \det (\nabla h(x) \nabla h(x)^\top)^{-1/2} \lambda_{\mathcal{M}}(dx)} - P^*\{\bar{B}\} \right| \]
\[
\left| \frac{\int_{\bar{B}} a(x) \det \left( \nabla h(x) \nabla h(x)^\top \right)^{-1/2} \lambda_M(dx)}{\int_{K_\eta} a(x) \det \left( \nabla h(x) \nabla h(x)^\top \right)^{-1/2} \lambda_M(dx)} - \frac{P_{\varepsilon_k} \{\bar{B}\}}{P_{\varepsilon_k} \{K_\eta\}} \right| \\
+ \left| \frac{P_{\varepsilon_k} \{\bar{B}\}}{P_{\varepsilon_k} \{K_\eta\}} - \bar{B} \right| - \left| \frac{P_{\varepsilon_k} \{\bar{B}\} - P^* \{\bar{B}\}}{P_{\varepsilon_k} \{K_\eta\}} \right|.
\]

(51)

In \eqref{51}, (I) can be made arbitrarily small, say \((I) \leq \eta\), for sufficiently large \(k\) as \(\mathcal{M} \cap K_\eta\) is a compact manifold with boundary and the result having been proved in Section A.1 applies. (II) \(\leq \eta P_{\varepsilon_k} \{\bar{B}\}/(1 - \eta) \leq 2\eta\) for all \(\eta \in (0, 1/2)\). (III) \(\leq \eta\) for sufficiently large \(k\) due to the weak convergence of the subsequence and the Portmanteau Lemma. Altogether \(\leq 4\eta\) when \(k\) is sufficiently large. The left-hand side of \eqref{51} then vanishes as we send \(\eta\) to 0. It follows that \(\int a(x) \det \left( \nabla h(x) \nabla h(x)^\top \right)^{-1/2} \lambda_M(dx) = \lim_{\eta \downarrow 0} \int_{K_\eta} a(x) \det \left( \nabla h(x) \nabla h(x)^\top \right)^{-1/2} \lambda_M(dx) < \infty\).

The remaining task is to show that \(P_0\) and \(P^*\) coincide, which implies that the full sequence \(P_\varepsilon \Rightarrow P_0\) as the converging subsequence \(\{P_{\varepsilon_k}\}\) is arbitrarily chosen. Let \(\mathcal{P} = \{\text{all bounded } P^*-\text{continuity sets}\}\) — \(\mathcal{P}\) is closed under finite intersections and thus a \(\pi\)-system. For reasons that \(\mathcal{P}\) contains all the \(P^*-\text{continuity } \ell_2\)-balls and that at most countably many open balls centered at each point \(x \in \mathcal{R}^d\) are not \(P^*-\text{continuity sets}, \mathcal{P}\) generates the Borel \(\sigma\)-algebra in \(\mathcal{R}^d\). Because \(\ell_2\)-balls are bounded, \(P_0 = P^*\) on \(\mathcal{P}\) by \eqref{51}; Dynkin’s \(\pi\)-\(\lambda\) Theorem therefore extends the equity of \(P_0\) and \(P^*\) on all Borel sets. We now conclude the proof of Theorem.\[1\]
Appendix B

Proof of Lemma 2

Under Assumption 1, we have \( \nabla \theta G(u, \theta)^\top [G(u, \theta) - y] = 0 \) for all \((u^\top, \theta^\top)^\top \in G(y)\), which is the first-order condition for minimizing \( \|G(u, \theta) - y\|^2/2 \) with respect to \( \theta \). By Assumption 1 and the Implicit Function Theorem (Rudin 1964, Chapter 9), there exists a unique, twice continuously differentiable function \( \hat{\theta}(y, u) \) such that

\[
\nabla \theta G(u, \hat{\theta}(y, u))^\top [G(u, \hat{\theta}(y, u)) - y] = 0
\]

for all \( u \) in some neighborhood of \( U(y) \). The isomorphism between \( G(y) \) and \( U(y) \) follows from the uniqueness of \( \hat{\theta}(y, u) \) for all \( u \in U(y) \).

Project \( G(u, \theta) - y \) to the range and null spaces of \( \nabla \theta G(u, \theta) \), respectively:

\[
G(u, \theta) - y = \nabla \theta G(u, \theta) \left( \nabla \theta G(u, \theta)^\top \nabla \theta G(u, \theta) \right)^{-1} \nabla \theta G(u, \theta)^\top [G(u, \theta) - y] \\
+ \nabla \theta \overline{G}(u, \theta) \nabla \theta \overline{G}(u, \theta)^\top [G(u, \theta) - y].
\]

Upon replacing \( \theta \) by \( \hat{\theta}(y, u) \), the first term on the right-hand side of (53) vanishes due to (52). Hence, \( G(u, \hat{\theta}(y, u)) - y \) for \( u \in U(y) \) reduces to

\[
h(u) = \nabla \theta \overline{G}(u, \hat{\theta}(y, u)) \nabla \theta \overline{G}(u, \hat{\theta}(y, u))^\top [G(u, \hat{\theta}(y, u)) - y],
\]

which justifies that \( U(y) \) can be implicitly defined as the level set \( h(u) = 0 \), i.e. (21). \( h(u) \) is twice continuously differentiable, which follows from the three-time continuous differentiability of \( G \) and the full-rank assumption for \( \nabla \theta G \). Moreover, note that

\[
\nabla h(u) = \nabla \theta \overline{G}(u, \hat{\theta}(y, u))^\top \nabla \theta G(u, \hat{\theta}(y, u))
\]

for all \( u \in U(y) \). As \( \nabla \theta G(u, \hat{\theta}(y, u)) \) is assumed to have a full row rank, the rank of \( \nabla h(u) \) equals to \( n - q \), and thus \( h(u) \) is a submersion.

Finally, we establish (23) that translates between the intrinsic measures \( \lambda_{G(y)} \) and
$\lambda_{U(y)}$. Let the $m \times (m - n + q)$ matrix $u(\omega)$ denote a smooth local parameterization of the manifold $U(y)$ (see Remark 1). We then have

$$
\lambda_{U(y)}(du) = \det \left( \nabla u(\omega)^\top \nabla u(\omega) \right)^{1/2} d\omega \\
= \det \left( \nabla u(\omega)^\top \nabla u(\omega) \right)^{1/2} \frac{\det \left( \nabla u(\omega)^\top \nabla u(\omega) + \nabla u(\omega)^\top \nabla_u \hat{\theta}(y, u(\omega)) \nabla_u \hat{\theta}(y, u(\omega)) \nabla u(\omega) \right)^{1/2}}{\det \left( \nabla u(\omega)^\top \nabla u(\omega) \right)^{1/2}} \lambda_{G(y)}(du, d\theta).
$$

(56)

In (56), we have

$$
\nabla_u \hat{\theta}(y, u) = - \left[ \nabla_\theta G(u, \hat{\theta}(y, u)) \right]^{-1} \nabla_\theta G(u, \hat{\theta}(y, u)) \nabla_u G(u, \hat{\theta}(y, u)),
$$

(57)

for $u \in U(y)$, which is obtained by differentiating (52) with respect to $u$.

Our remaining task is to show that the ratio in the last line of (56) equals to $D(u, \theta)^{-1/2}$. Dependencies on $u$ and $\theta$ are dropped for the rest of the appendix to simplify the notation. By the Matrix Determinant Lemma and (57),

$$
\det \left( \nabla u^\top \nabla u + \nabla u^\top \nabla_u \hat{\theta}^\top \nabla_u \nabla u \right) \\
= \det (\nabla u^\top \nabla u) \cdot \det \left( \iota_q + \nabla_u \hat{\theta} \nabla u (\nabla u^\top \nabla u)^{-1} \nabla u^\top \nabla_u \hat{\theta}^\top \right) = \det (\nabla u^\top \nabla u) \\
\cdot \det \left( \iota_q + (\nabla_\theta G^\top \nabla_\theta G)^{-1} \nabla_\theta G^\top \nabla u G \nabla u (\nabla_u^\top \nabla u)^{-1} \nabla u^\top \nabla_u G^\top \nabla_\theta G (\nabla_\theta G^\top \nabla_\theta G)^{-1} \right).
$$

(58)

It further suffices to show that $\hat{D} = D$. Note that $\hat{D}$ depends on $\nabla u$ only through the projection matrix $\nabla u(\nabla u^\top \nabla u)^{-1} \nabla u^\top$, so $\hat{D}$ is invariant to different choices of $\nabla u$. We proceed to set

$$
\nabla u = (K : L),
$$

(59)

31
in which $K$ is an $m \times (m - n)$ orthonormal complement of $\nabla_u G^\top$, i.e., $K^\top \nabla_u G^\top = 0$, and $L = \nabla_u G^\top (\nabla_u G \nabla_u G^\top)^{-1} \nabla_\theta G$ with dimension $m \times q$. To see that (59) is a valid choice of $\nabla u$, we first note that $(K : L)$ has a full column rank equal to $m - n + q$: This is because $L$ is of full column rank by Assumption 1(ii) and $K^\top L = 0$. Moreover,

$$\nabla_\theta G^\top \nabla_u G (K : L) = \nabla_\theta G^\top (0 : \nabla_\theta G) = 0,$$

so $(K : L)$ is perpendicular to the $m \times (n - q)$ matrix $\nabla_u G^\top \nabla_\theta G$ that spans the normal space of $U(y)$. Our choice of $\nabla u$ allows us to simplify $\tilde{D}$ to

$$\tilde{D} = \det \left(t_q + (L^\top L)^{-1}\right) = \det \left(t_q + [\nabla_\theta G^\top (\nabla_u G \nabla_u G^\top)^{-1} \nabla_\theta G]^{-1} \right) = D. \quad (61)$$

The proof of the lemma is complete.

\footnote{Assumption 1(ii) implies that $m \geq n$; if $m = n$, simply remove $K$ from (59).}
Appendix C

Proof of Propositions 4 and 5

C.1 Proof of Proposition 4

(27) is a direct consequence of Theorem 1. To establish (28), we apply the Matrix Determinant Lemma twice to (24):

\[ D = \frac{\det \left( \iota_q + \nabla_\theta G^\top (\nabla_u G \nabla_u G^\top)^{-1} \nabla_\theta G \right)}{\det \left( \nabla_\theta G^\top (\nabla_u G \nabla_u G^\top)^{-1} \nabla_\theta G \right)} = \frac{\det \left( \nabla_u G \nabla_u G^\top + \nabla_\theta G \nabla_\theta G^\top \right)}{\det (\nabla_u G \nabla_u G^\top)} \det \left( \nabla_\theta G^\top (\nabla_u G \nabla_u G^\top)^{-1} \nabla_\theta G \right). \tag{62} \]

(28) immediately follows from (62) and the change of measure in Lemma 2.

To arrive at the last statement of Proposition 4, we apply the smooth coarea formula (Chavel, 2006, Section III.8): For any measurable \( B \subseteq \Theta \),

\[ \int_{(y,u) \in B} \tilde{f}_B(u) \lambda_{U(y)}(du) = \int_B \left[ \int \frac{\tilde{f}_B(u)}{\det \left( \nabla_u \hat{\theta}_{|T_u U(y)}^\top \nabla_u \hat{\theta}_{|T_u U(y)} \right)^{1/2} \lambda_{\theta(y)}} \right] d\theta, \tag{63} \]

in which \( \nabla_u \hat{\theta}_{|T_u U(y)} \) refers to the \((m - n + q) \times q\) matrix obtained by projecting the \(m \times q\) dimensional \( \nabla_u \hat{\theta} \) to the \((m - n + q)\)-dimensional tangent space of \( U(y) \) at \( u \). Choosing (59) as the basis for \( T_u U(y) \), we have

\[ \det \left( \nabla_u \hat{\theta}_{|T_u U(y)}^\top \nabla_u \hat{\theta}_{|T_u U(y)} \right) = \det \left( \nabla_\theta G^\top (\nabla_u G \nabla_u G^\top)^{-1} \nabla_\theta G \right)^{-1}. \tag{64} \]

Therefore, the bracketed term on the right-hand side of (63) can be further written as

\[ \int \frac{\tilde{f}_B(u)}{\det \left( \nabla_u \hat{\theta}_{|T_u U(y)}^\top \nabla_u \hat{\theta}_{|T_u U(y)} \right)^{1/2} \lambda_{\theta(y)}} \propto \pi(\theta) \left[ \int \frac{\rho(u)}{\det \left( \nabla_u G \nabla_u G^\top \right)^{1/2} \lambda_{\theta(y)}} \right], \tag{65} \]

which is proportional to the posterior density due to (25).
C.2 Proof of Proposition 5

(29) is a direct consequence of Theorem 1. To establish (30), we derive an expression for \( \det \left( \nabla_\theta G^\top \nabla_u G \nabla_u G^\top \nabla_\theta G \right) \) that does not explicitly involve the orthogonal complement \( \nabla_\theta G \). Let \( A = (\nabla_\theta G : \nabla_\theta G) \), which is a full-rank square matrix with dimension \( n \times n \). Applying the Schur Determinant Identity to \( A^\top \nabla_u G \nabla_u G^\top A \), we obtain

\[
\begin{align*}
\det (A^\top \nabla_u G \nabla_u G^\top A) &= \det(A)^2 \det (\nabla_u G \nabla_u G^\top) = \det(\nabla_\theta G^\top \nabla_u G \nabla_u G^\top \nabla_\theta G) \\
&\cdot \det \left( \nabla_\theta G^\top \nabla_u G \left[ I_{n \times n} - \nabla_u G^\top \nabla_\theta G (\nabla_\theta G^\top \nabla_u G \nabla_u G^\top \nabla_\theta G)^{-1} \nabla_\theta G^\top \nabla_u G \right] \nabla_u G^\top \nabla_\theta G \right) \\
&= \det \left( \nabla_\theta G^\top \nabla_u G \nabla_u G^\top \nabla_\theta G \right) \det \left( \nabla_\theta G^\top \nabla_u G \nabla_u G^\top \nabla_\theta G \right)^2 \\
&= \frac{\det \left( \nabla_\theta G^\top (\nabla_u G \nabla_u G^\top)^{-1} \nabla_\theta G \right)}{\det \left( \nabla_\theta G^\top \nabla_\theta G \right)} ,
\end{align*}
\]

which follows from (59) and (60). (66) and the equality \( \det(A)^2 = \det(\nabla_\theta G^\top \nabla_\theta G) \) imply that

\[
\begin{align*}
\det \left( \nabla_\theta G^\top \nabla_u G \nabla_u G^\top \nabla_\theta G \right) &= \frac{\det (\nabla_u G \nabla_u G^\top) \det \left( \nabla_\theta G^\top (\nabla_u G \nabla_u G^\top)^{-1} \nabla_\theta G \right)}{\det (\nabla_\theta G^\top \nabla_\theta G)} .
\end{align*}
\]

(30) is then deduced from (62), (67), and Lemma 2.

Similar to (63)–(65), the last statement of Proposition 5 also follows directly from the smooth coarea formula:

\[
\int_{\hat{\theta}(y,u) \in B} \tilde{f}_F(u) \lambda_{U(y)}(du) = \int_B \left[ \int_B \frac{\tilde{f}_F(u)}{\det \left( \nabla_u \hat{\theta} |_{T_u U(y)} \nabla_u \hat{\theta} |_{T_u U(y)} \right)^{1/2} \lambda_{G(\theta)}(\theta)} \right] d\theta .
\]

The bracketed term in (68) gives the density of \( \hat{\theta}(y,u) \), which is proportional to (31) as a result of (67).
Appendix D

Proof of Proposition 6

For succinctness, we treat \( E \) as an \( IJ \times 1 \) vector throughout this proof, replacing the notation \( \text{vec}(E) \) in the main text. The repeated-measures ANOVA model can then be expressed in matrix form as

\[
Y = (1_J \otimes \iota_I) \mu + \sigma_z (\iota_J \otimes 1_I) Z + \sigma_e E
\]

\[
= W(Z) \beta + \sigma_e E. \tag{69}
\]

in which \(^{10}W(z) = (1_J \otimes \iota_I : (\iota_J \otimes 1_I) z) \in \mathcal{R}^{I+1} \) and \( \beta = (\mu^\top, \sigma_z)^\top \). Also let \( W(z) \in \mathcal{R}^{IJ-I-1} \) be an orthonormal complement of \( W(z) \), and \( r(y, z) = W(z) W(z)^\top y \) be the projection of \( y \) onto the null space of \( W(z) \)—equivalently, the residual after regressing \( y \) on \( W(z) \).

We proceed to characterize the set

\[
C_\varepsilon(y, z) = \{ e \in \mathcal{R}^{IJ} : \min_{\beta, \sigma_e} \| W(z) \beta + \sigma_e e - y \| \leq \varepsilon \}, \tag{70}
\]

There are two cases to consider. First, if \( \| r(y, z) \| \leq \varepsilon \), then \( C_\varepsilon(y, z) = \mathcal{R}^{IJ} \) because the minimum of \( \| W(z) \beta - y \| \) (i.e., fixing \( \sigma_e \) at 0) is already no greater than \( \varepsilon \). Second, if \( \| r(y, z) \| > \varepsilon \), then the least-square solution of \( \sigma_e \) (i.e., \( \hat{\sigma}_e \)) corresponding to \( e \in C_\varepsilon(y, z) \) must be non-zero. We claim that \( C_\varepsilon(y, z) \) in the second case is equivalent to

\[
\tilde{C}_\varepsilon(y, z) = \{ [r(y, z) + e_1] \alpha + e_2 : e_1 \in \mathfrak{n}(W(z)), \| e_1 \| \leq \varepsilon, \alpha \neq 0, e_2 \in \mathfrak{r}(W(z)) \}, \tag{71}
\]

in which \( \mathfrak{r}(A) \) and \( \mathfrak{n}(A) \) denote the range and null space for the columns of \( A \). To see this, first take \( e \in C_\varepsilon(y, z) \), which satisfies \( y = W(z) \hat{\beta} + \hat{\sigma}_e e + \varrho \) with some \( \varrho \in \mathcal{R}^{IJ} \) such that \(^{10}W(z) \) is rank deficient (with rank \( I < I + 1 \)) when \( z \) is a multiple of \( 1_J \).
\[ \|\varrho\| \leq \varepsilon. \]  Because \( \hat{\sigma}_e \neq 0 \), we have

\[ e = \mathbf{W}(z)\mathbf{W}(z)^\top \left( \frac{y - \varrho}{\hat{\sigma}_e} \right) + \left[ \iota_{IJ} - \mathbf{W}(z)\mathbf{W}(z)^\top \right] \left( \frac{y - \varrho}{\hat{\sigma}_e} \right) - \frac{W(z)\hat{\beta}}{\hat{\sigma}_e} \]

\[ = [r(y, z) - \mathbf{W}(z)\mathbf{W}(z)^\top \varrho] \sigma_e^{-1} + \left[ \iota_{IJ} - \mathbf{W}(z)\mathbf{W}(z)^\top \right] \left( \frac{y - \varrho}{\hat{\sigma}_e} \right) - \frac{W(z)\hat{\beta}}{\hat{\sigma}_e}, \quad (72) \]

which can be identified as an element in \( \tilde{C}_\varepsilon(y, z) \). Conversely, take \( e \in \tilde{C}_\varepsilon(y, z) \) so that

\[ e = (r(y, z) + e_1)\alpha + W(z)\beta_1 \]

for some \( \alpha \neq 0, \beta_1 \in \mathcal{R}^{I+1} \), and \( e_1 \in \mathfrak{n}(W(z)) \) such that \( \|e_1\| \leq \varepsilon \). As \( r(y, z) = y - W(z)\beta_2 \) for some \( \beta_2 \in \mathcal{R}^{I+1} \), we let \( \beta = \beta_1 - \beta_2\alpha \) and therefore have

\[ \|y + \frac{W(z)\beta}{\alpha} - \frac{e}{\alpha}\| = \| - e_1\| \leq \varepsilon, \quad (73) \]

which implies \( e \in C_\varepsilon(y, z) \). Geometrically, \( \tilde{C}_\varepsilon(y, z) \) in (71) is the Cartesian product of \( r(W(z)) \) and a double (spherical) cone in \( \mathfrak{n}(W(z)) \) centered at the origin. This is because in (71) \( r(y, z) + e_1 \) with \( \|e_1\| \leq \varepsilon \) falls within an \( \ell_2 \)-ball around \( r(y, z) \) with radius \( \varepsilon < r(y, z) \); therefore, points that are multiples of \( r(y, z) + e_1 \) form a spherical cone with central angle \( \sin^{-1}(\varepsilon/\|r(y, z)\|) \).

Our final task is to find compact sets \( K \in \mathcal{R}^J \) and \( L \in \mathcal{R}^{IJ} \) such that the ratio

\[ \frac{\mathbb{P}\{\{E \in C_\varepsilon(y, Z) \cap L\} \cap \{Z \in K\}\}}{\mathbb{P}\{E \in C_\varepsilon(y, Z)\}} \quad (74) \]

can be made arbitrarily close to 1. Taking advantage of the spherical symmetry in our setup, let \( K \in \mathcal{R}^J \) and \( L \in \mathcal{R}^{IJ} \) be closed \( \ell_2 \)-balls centered at the origin. Because \( E \) follows a spherical distribution independent of \( Z \) and \( C_\varepsilon(y, z) \) is spherically symmetric,

\[ \mathbb{P}\{E \in C_\varepsilon(y, z) \cap L \mid Z = z\} = \mathbb{P}\{E \in L\} \cdot \varpi(y, z, \varepsilon), \quad (75) \]

where \( \varpi(y, z, \varepsilon) \in (0, 1] \) is equal to 1 if \( \|r(y, z)\| \leq \varepsilon \) and otherwise is a monotonically increasing function of the spherical cone’s central angle. Because \( Z \) also follows a spherical distribution, let \( Z = R_z \mathbf{V}_z \) where \( \mathbf{V}_z \in \mathcal{R}^J \) is uniform on the unit sphere and \( R_z > 0 \) is
independent of $V_z$. The earlier geometric analysis reveals that the central angle of the spherical cone depends only on $V_z$ but not $R_z$. It follows that

$$
\mathbb{P}\left\{\{E \in C_{\varepsilon}(y, Z) \cap L\} \cap \{Z \in K\}\right\} = \int_K \mathbb{P}\left\{E \in C_{\varepsilon}(y, z) \cap L \mid Z = z\right\} \mathbb{P}(dz)
$$

$$
= \mathbb{P}\{E \in L\} \int_K \varpi(y, z, \varepsilon) \mathbb{P}(dz) = \mathbb{P}\{E \in L\} \int \varpi(y, v_z, \varepsilon) d\mathbb{P}(dv_z) \int_0^{\gamma(K)} d\mathbb{P}(dr_z), \quad (76)
$$

in which $v_z$ and $r_z$ are respective realizations of $V_z$ and $R_z$, and $\gamma(K)$ denotes the radius of $K$. Similarly,

$$
\mathbb{P}\{E \in C_{\varepsilon}(y, Z)\} = \int \mathbb{P}\{E \in C_{\varepsilon}(y, z) \mid Z = z\} \mathbb{P}(dz)
$$

$$
= \int \varpi(y, z, \varepsilon) \mathbb{P}(dz) = \int \varpi(y, v_z, \varepsilon) d\mathbb{P}(dv_z). \quad (77)
$$

Hence, the ratio of (76) over (77) is

$$
\frac{\mathbb{P}\left\{\{E \in C_{\varepsilon}(y, Z) \cap L\} \cap \{Z \in K\}\right\}}{\mathbb{P}\{E \in C_{\varepsilon}(y, Z)\}} = \frac{\mathbb{P}\{E \in L\} \mathbb{P}\{Z \in K\}}{\mathbb{P}\{E \in C_{\varepsilon}(y, Z)\}}, \quad (78)
$$

which is constant in $\varepsilon$ and can be made arbitrarily close to 1.
Appendix E
Computational Complexity for Repeated-Measures ANOVA

E.1 Evaluating the Fiducial Density

When \( I, J > 1, \) \( IJ \geq I + 2. \) The matrix\(^{11}\) \( \nabla_u G \nabla_u G^\top + \nabla_\theta G \nabla_\theta G^\top \) is then a low-rank modification to the matrix \( \nabla_u G \nabla_u G^\top, \) which allows us to use well-known linear algebraic results such as the Woodbury identity and Matrix Determinant Lemma to lessen the computational burden.

Because the schoolbook complexity for computing \( \det(\nabla_\theta G^\top \nabla_\theta G), \) which appears as the first determinant term on the right-hand side of (30), is already \( O(I^3J) \), we focus on the second determinant term, i.e., (32) after applying the Matrix Determinant Lemma. Note that

\[
\nabla_u G \nabla_u G^\top = I_J \otimes \left( \sigma^2_e 1_I + \sigma^2_z 1_I 1_I^\top \right).
\]

(79) has a repetitive block-diagonal structure. The \( I \times I \) diagonal block \( \Omega = \sigma^2_e 1_I + \sigma^2_z 1_I 1_I^\top \) is a rank-one modification to a diagonal matrix: \( \Omega^{-1} = \sigma^{-2} e 1_I - \sigma^{-4} \sigma^2_z (1 + I \sigma^{-2} \sigma^2_z)^{-1} 1_I 1_I^\top \) by the Woodbury formula and \( \det(\Omega) = \sigma^{2I} (1 + I \sigma^{-2} \sigma^2_z) \) by the Matrix Determinant Lemma. Therefore, solving the linear system \( (I_J \otimes \Omega)x = b \) for \( x, b \in R^{IJ} \) takes only \( O(IJ) \) flops rather than \( O(I^3J^3) \) flops that would have been needed for an unstructured left-hand side matrix. This further reduces the computation of \( \nabla_\theta G (\nabla_u G \nabla_u G^\top)^{-1} \nabla_\theta G \) to \( O(I^3J) \) flops assuming the schoolbook complexity for matrix multiplication and determinant calculation. Because the second determinant in (32) takes also \( O(I^3J) \) flops to evaluate, the overall complexity of evaluating the fiducial density is \( O(I^3J) \) rather than \( O(I^3J^3). \)

E.2 Manifold MCMC Update

The complexity of the manifold RWM/HMC update is determined by two operations: finding an orthonormal basis for the null space of \((\nabla_u G : \nabla_\theta G)^\top\) and retracting a point back to the manifold (i.e., Algorithm 2). The retraction step solves linear equations with left-hand side matrices of the form \( \nabla_u G(u, \theta) \nabla_u G(u', \theta')^\top + \nabla_\theta G(u, \theta) \nabla_\theta G(u', \theta')^\top, \)

\(^{11}\) For notational succinctness, we again suppress the dependency on \( u \) and \( \theta. \)
where \( u, u' \in \mathcal{R}^{(I+1)J}, \theta, \theta' \in \mathcal{R}^{I+2} \): An argument similar to the previous paragraph shows that its complexity is \( O(I^3J) \).

An orthonormal basis matrix of the null space is routinely obtained via a full QR decomposition: For \( (\nabla_u G : \nabla_\theta G)^\top \), it takes \( O(I^3J^3) \) flops. Nevertheless, we can take advantage of the fact that \( \nabla_u G \) contains a \( IJ \times IJ \) diagonal block. In particular, it can be straightforwardly verified that

\[
\begin{pmatrix}
-\sigma e^I_{I+2} \\
\iota_J \otimes \sigma_z 1_I : \nabla_\theta G
\end{pmatrix}
\tag{80}
\]

is an orthogonal complement of \( (\iota_J \otimes \sigma_z 1_I : \nabla_\theta G : \sigma e^I_{I+2})^\top \), which becomes \( (\nabla_u G : \nabla_\theta G)^\top \) after a suitable permutation of rows. The remaining task is to orthogonalizing and normalizing the columns of \([80]\), which amounts to QR-factorizing \( (\iota_J \otimes \sigma_z 1_I : \nabla_\theta G) \) because the diagonal block \( -\sigma e^I_{I+2} \) already has orthogonal columns. Note that the projection matrix corresponding to \( \iota_J \otimes \sigma_z 1_I \) is \( \iota_J \otimes I^{-1}1_I 1_I^\top \), which again has a repetitive block-diagonal structure. It then suffices to first project \( \nabla_\theta G \) to the null space of \( \iota_J \otimes \sigma_z 1_I \) and then apply a QR decomposition, each of which takes only \( O(I^3J) \) flops.
References

Alquier, P., N. Friel, R. Everitt, and A. Boland (2016). Noisy Monte Carlo: Convergence of markov chains with approximate transition kernels. *Statistics and Computing* 26(1), 29–47.

Beaumont, M. A. (2019). Approximate Bayesian computation. *Annual Review of Statistics and Its Application* 6, 379–403.

Beaumont, M. A., W. Zhang, and D. J. Balding (2002). Approximate Bayesian computation in population genetics. *Genetics* 162(4), 2025–2035.

Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian analysis* 1(3), 385–402.

Berger, J. O., J. M. Bernardo, and D. Sun (2015). Overall objective priors. *Bayesian Analysis* 10(1), 189–221.

Brubaker, M., M. Salzmann, and R. Urtasun (2012). A family of MCMC methods on implicitly defined manifolds. In *Artificial Intelligence and Statistics*, pp. 161–172. PMLR.

Chavel, I. (2006). *Riemannian Geometry: A Modern Introduction*. (2nd ed.). Cambridge University Press.

Cisewski, J. and J. Hannig (2012). Generalized fiducial inference for normal linear mixed models. *The Annals of Statistics* 40(4), 2102–2127.

Cranmer, K., J. Brehmer, and G. Louppe (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences* 117(48), 30055–30062.

Dempster, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *The Annals of Mathematical Statistics*, 355–374.

Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)* 30(2), 205–232.
Dempster, A. P. (2008). The Dempster–Shafer calculus for statisticians. *International Journal of approximate reasoning* 48(2), 365–377.

Fang, K., S. Kotz, and K. Ng (1990). *Symmetric Multivariate and Related Distributions*. Taylor & Francis.

Fisher, R. A. (1925). Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* 22(5), 700–725.

Fisher, R. A. (1930). Inverse probability. *Mathematical Proceedings of the Cambridge Philosophical Society* 26(4), 528–535.

Fisher, R. A. (1933). The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proceedings of the Royal Society of London, Series A* 139(838), 343–348.

Fisher, R. A. (1935). The fiducial argument in statistical inference. *Annals of Eugenics* 6(4), 391–398.

Fraser, D. A. S. and K. W. Ng (1980). Multivariate regression analysis with spherical error. *Multivariate analysis* 5, 369–386.

Fraser, D. A. S., N. Reid, E. Marras, and G. Y. Yi (2010). Default priors for Bayesian and frequentist inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(5), 631–654.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2013). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: Taylor & Francis.

Hannig, J. (2009). On generalized fiducial inference. *Statistica Sinica*, 491–544.

Hannig, J. (2013). Generalized fiducial inference via discretization. *Statistica Sinica*, 489–514.
Hannig, J., H. Iyer, R. C. Lai, and T. C. Lee (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association* 111(515), 1346–1361.

Hwang, C.-R. (1980). Laplace’s method revisited: Weak convergence of probability measures. *The Annals of Probability*, 1177–1182.

Kass, R. E. and L. Wasserman (1996). The selection of prior distributions by formal rules. *Journal of the American statistical Association* 91(435), 1343–1370.

Lai, R. C. S., J. Hannig, and T. C. M. Lee (2015). Generalized fiducial inference for ultrahigh-dimensional regression. *Journal of the American Statistical Association* 110(510), 760–772.

Lee, J. (2013). *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer New York.

Lelièvre, T., M. Rousset, and G. Stoltz (2012). Langevin dynamics with constraints and computation of free energy differences. *Mathematics of computation* 81(280), 2071–2125.

Lelièvre, T., M. Rousset, and G. Stoltz (2019). Hybrid Monte Carlo methods for sampling probability measures on submanifolds. *Numerische Mathematik* 143(2), 379–421.

Lelièvre, T., G. Stoltz, and W. Zhang (2022). Multiple projection Markov chain Monte Carlo algorithms on submanifolds. *IMA Journal of Numerical Analysis*.

Liu, Y. and J. Hannig (2016). Generalized fiducial inference for binary logistic item response models. *Psychometrika* 81(2), 290–324.

Liu, Y. and J. Hannig (2017). Generalized fiducial inference for logistic graded response models. *Psychometrika* 82(4), 1097–1125.

Marin, J.-M., P. Pudlo, C. P. Robert, and R. J. Ryder (2012). Approximate Bayesian computational methods. *Statistics and Computing* 22(6), 1167–1180.
Martin, R. and C. Liu (2013). Inferential models: A framework for prior-free posterior probabilistic inference. *Journal of the American Statistical Association* **108**(501), 301–313.

Martin, R. and C. Liu (2015a). Conditional inferential models: Combining information for prior-free probabilistic inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**(1), 195–217.

Martin, R. and C. Liu (2015b). *Inferential models: Reasoning with uncertainty*. CRC Press.

Martin, R. and C. Liu (2015c). Marginal inferential models: Prior-free probabilistic inference on interest parameters. *Journal of the American Statistical Association* **110**(512), 1621–1631.

MATLAB (2021). *version 9.11.0 (R2021b)*. Natick, Massachusetts: The MathWorks Inc.

Mukerjee, R. and D. K. Dey (1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter: Higher order asymptotics. *Biometrika* **80**(3), 499–505.

Mukerjee, R. and M. Ghosh (1997). Second-order probability matching priors. *Biometrika* **84**(4), 970–975.

Murph, A., J. Hannig, and J. P. Williams (2022a). Introduction to generalized fiducial inference. In S. Brooks, A. Gelman, G. Jones, and X.-L. Meng (Eds.), *Handbook on Bayesian, Frequentist, and Fiducial Inference*. Chapman & Hall.

Murph, A. C., J. Hannig, and J. P. Williams (2022b). Generalized fiducial inference on differentiable manifolds. https://arxiv.org/abs/2209.15473.

Potthoff, R. F. and S. Roy (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**(3-4), 313–326.

Rudin, W. (1964). *Principles of Mathematical Analysis*. McGraw-Hill.
Schweder, T. and N. L. Hjort (2002). Confidence and likelihood. Scandinavian Journal of Statistics 29(2), 309–332.

Shi, J., J. Hannig, R. C. S. Lai, and T. C. M. Lee (2021). Covariance estimation via fiducial inference. Statistical Theory and Related Fields 5(4), 316–331.

Singh, K., M.-G. Xie, and W. E. Strawderman (2005). Combining information from independent sources through confidence distributions. The Annals of Statistics 33(1), 159–183.

Welch, B. L. and H. W. Peers (1963). On formulae for confidence points based on integrals of weighted likelihoods. Journal of the Royal Statistical Society: Series B (Methodological) 25(2), 318–329.

Weyl, H. (1939). On the volume of tubes. American Journal of Mathematics 61(2), 461–472.

Xie, M.-G. and K. Singh (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. International Statistical Review 81(1), 3–39.

Xie, M.-G., K. Singh, and W. E. Strawderman (2011). Confidence distributions and a unifying framework for meta-analysis. Journal of the American Statistical Association 106(493), 320–333.

Zappa, E., M. Holmes-Cerfon, and J. Goodman (2018). Monte Carlo on manifolds: Sampling densities and integrating functions. Communications on Pure and Applied Mathematics 71(12), 2609–2647.