Fiducial theory for free-knot splines

Derek L. Sonderegger[†,*] **and Jan Hannig**[‡,§]

Abstract: We construct the fiducial model for free-knot splines and derive sufficient conditions to show asymptotic consistency of a multivariate fiducial estimator. We show that splines of degree four and higher satisfy those conditions and conduct a simulation study to evaluate quality of the fiducial estimates compared to the competing Bayesian solution. The fiducial confidence intervals achieve the desired confidence level while tending to be shorter than the corresponding Bayesian credible interval using the reference prior.

AMS 2000 subject classifications: Primary 62F99, 62G08; secondary 62P10.
Keywords and phrases: free-knot splines, generalized fiducial inference, Bernstein-von Mises theorem.

1. Introduction

In statistical practice there is a tension between fitting an easily interpretable model to our data versus fitting a highly flexible model that fits the data better. One compromise between these competing ideas is a spline model. The spline model of degree p can be thought of as connected degree p polynomials with the requirement that the resulting function be "smooth" at the connection points. These connection points are usually called "knot points" and the usual smoothness requirement is that the $p-1$ derivative exists.

The simplest example is the $p = 1$ spline with 1 knot point, which is a linear function with some slope until the knot point, and then continues with a different slope. The smoothness requirement is that the 0th derivative exists, which is that the function is continuous at the knot point. The resulting function is often called the hockey-stick function. A degree $p = 2$ spline with 1 knot point is just two quadratic curves joined together such that at the knot point the function has a 1st derivative and is therefore "smooth".

When using splines to approximate an unknown but continuous function, one important question is where to place the knots. In typical non-parametric function estimation, more knots than are necessary are evenly spread along the dependent axis and a penalty based on the second derivative (a.k.a. function "wiggliness") is introduced [13]. An alternative approach is to use a small number of knots but carefully place them. This problem of where to place the knots

*Corresponding Author: derek.sonderegger@nau.edu
†Department of Mathematics and Statistics, Northern Arizona Univeristy
‡Department of Statistics and Operations Research, University of North Carolina at Chapel Hill
§Jan Hannig's research was supported in part by the National Science Foundation under Grant No. 1007543 and 1016441.

is known as the free-knot spline problem. The free-knot spline problem is primarily interested in estimating the location of the knot point and interpreting it as some sort of threshold [15, 14].

A Bayesian solution to the arbitrary degree p problem with a fixed number of knot points is given by DiMatteo, Genovese and Kass [2] and they recommend using a prior of $p\left(\boldsymbol{\alpha}, \boldsymbol{t}, \sigma^2\right) \propto \sigma^{-2}$ where σ^2 is the usual variance term, $\boldsymbol{\alpha}$ is the polynomial coefficients, and $\boldsymbol{t}$ is the vector of knot points. The maximum likelihood solution for the degree $p = 1$ free-knot spline problem is developed in [11] and is available in the R package `segmented` [12].

In this paper, we investigate the fiducial solution to the free-knot spline problem of degree $p \geq 4$. In Section 2, we first extend the univariate fiducial Bernstein-von Mises theorem to the multivariate setting, which shows that multivariate fiducial estimators have an asymptotic multivariate normal distribution under certain assumptions. In Section 3, we derive the fiducial solution to the free-knot spline problem, note that the Bernstein-von Mises assumptions are satisfied and investigate the small sample properties by conducting a simulation study of degree $p = 4$ splines comparing the fiducial solution to the Bayesian solution of DiMatteo, Genovese and Kass [2]. In Section 4 we give our concluding remarks.

1.1. Introduction to fiducial inference

R.A. Fisher first introduced his idea of fiducial inference [4] to address what he felt was the major shortcoming of Bayesian inference. His goal was to invent a posterior-like distribution without the need for a prior distribution. He did not succeed in developing a general theory for finding these fiducial distributions and his idea was met with extreme skepticism. In the 1990's, generalized confidence intervals [19] were found to have very good small sample properties and Hannig, Iyer and Patterson [8] shows the connection between generalized confidence intervals and Fisher's fiducial inference. Hannig [6] develops a general theory for developing fiducial solutions which has been used in a variety of contexts. The solution for wavelets is given by [9]. Other problems include variance components in normal mixed linear model [3, 1], extreme value models [17], and multiple comparison issues [18].

The general framework of fiducial inference assumes that the n observed data can be written as a data generating equation $\mathbb{X} = \mathbf{G}(\mathbb{U}, \boldsymbol{\xi})$, where $\boldsymbol{\xi}$ is a p length vector of parameters, and $\mathbb{U}$ is a random vector of with a completely known distribution.

Setting $\mathbb{X}_0 = (X_1, \ldots, X_p)$, $\mathbb{X}_c = (X_{p+1}, \ldots, X_n)$, $\mathbb{U}_0 = (U_1, \ldots, U_p)$ and $\mathbb{U}_c = (U_{p+1}, \ldots, U_n)$ the data generating equation can be factorized as

$$\mathbb{X}_0 = \mathbf{G_0}(\mathbb{U}_0, \boldsymbol{\xi}) \quad \text{and} \quad \mathbb{X}_c = \mathbf{G_c}(\mathbb{U}_c, \boldsymbol{\xi}).$$

Assuming that for each $\boldsymbol{\xi} \in \boldsymbol{\Xi}$ that $\mathbf{G}_0(\boldsymbol{\xi}, \cdot)$ and $\mathbf{G}_c(\boldsymbol{\xi}, \cdot)$ are one-to-one and differentiable and that $\mathbf{G}_0(\boldsymbol{\xi}, \cdot)$ also invertible, then Hannig [6] shows that the

generalized fiducial distribution is

$$r(\boldsymbol{\xi}|\mathbf{x}_0) = \frac{f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\xi})J_0(\boldsymbol{x}_0,\boldsymbol{\xi})}{\int_{\boldsymbol{\Xi}} f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\xi}')J_0(\boldsymbol{x}_0,\boldsymbol{\xi}')d\boldsymbol{\xi}'}$$

where

$$J_0\left(\boldsymbol{x}_0,\boldsymbol{\xi}\right) = \left|\frac{det\left(\frac{d}{d\boldsymbol{\xi}}\mathbf{G}_0^{-1}\left(\boldsymbol{x}_0,\boldsymbol{\xi}\right)\right)}{det\left(\frac{d}{d\boldsymbol{x}_0}\mathbf{G}_0^{-1}\left(\boldsymbol{x}_0,\boldsymbol{\xi}\right)\right)}\right|$$

and $f_{\mathbf{x}}(x|\boldsymbol{\xi})$ is the density function. Since the choice to use the first p observations in the definition of $\mathbf{G}_0$ was arbitrary, we could select any p observations that satisfy the one-to-one, differentiable, and invertible conditions. Hannig suggests [6, 7] letting the Jacobian $J\left(\boldsymbol{x},\boldsymbol{\xi}\right)$ be the average of all possible values of J_0 and using

$$r(\boldsymbol{\xi}|\mathbf{x}) = \frac{f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\xi})J(\boldsymbol{x},\boldsymbol{\xi})}{\int_{\boldsymbol{\Xi}} f_{\mathbf{x}}(\boldsymbol{x}|\boldsymbol{\xi}')J(\boldsymbol{x},\boldsymbol{\xi}')d\boldsymbol{\xi}'}. \tag{1}$$

This distribution is similar to a Bayesian posterior distribution with the Jacobian taking the role of the prior. This can be seen in the standard regression problem where the Jacobian simplifies to $J\left(\boldsymbol{x},\boldsymbol{\xi}\right) = \sigma^{-2}h\left(\boldsymbol{x}\right)$. Since $h\left(\boldsymbol{x}\right)$ is in Jacobians in both the numerator and denominator, it will cancel and the fiducial distribution distribution is the same as the Bayesian posterior with commonly used reference prior distribution σ^{-2}.

Two numerical issues commonly arise in the evaluation of the fiducial density. First, it is often not feasible to take the average of all possible values of J_0 because the number of possible permutations grows as n^p. This is often solved by taking a random selection of possible J_0 and using the sample mean as an approximation to $J\left(\boldsymbol{x},\boldsymbol{\xi}\right)$. A second challenge comes in evaluating the denominator, which is often intractable due to the high number of dimensions. To address this issue, we use the standard Markov Chain Monte Carlo (MCMC) techniques to take a random sample from the fiducial density and all subsequent inference is based on that sample.

2. Asymptotic consistency of the multivariate fiducial estimators

Many estimators have an asymptotic normal distribution and fiducial estimators are no exception. Conditions A0-A6 in Appendix A are the standard conditions sufficient to prove that the maximum likelihood estimators to have an asymptotic normal distribution [10]. That is, the maximum likelihood estimators $\hat{\boldsymbol{\xi}}_n$ are consistent and $\sqrt{n}(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi})$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\left[I(\boldsymbol{\xi})\right]^{-1}$, where $I\left(\boldsymbol{\xi}\right)$ is the Fisher information matrix.

The Bernstein-von Mises theorem gives conditions (B1-B2 in appendix A.1) under which the Bayesian posterior distribution is asymptotically normal [16, 5]. In brief, the proof can be thought of as showing that the posterior distribution becomes close to the distribution of the MLE. Hannig [6] gives sufficient conditions (C1-C2) for the univariate fiducial distribution to converge to the Bayesian

posterior which is in turn is close the the MLE distribution. Hannig [6] defines the following assumptions:

(C1) For any $\delta > 0$

$$\inf_{\boldsymbol{\xi} \notin B(\boldsymbol{\xi}_0, \delta)} \frac{\min_{i=1\ldots n} \log f(\boldsymbol{\xi}, X_i)}{|L_n(\boldsymbol{\xi}) - L_n(\boldsymbol{\xi}_0)|} \xrightarrow{P_{\boldsymbol{\xi}_0}} 0$$

where $L_n(\boldsymbol{\xi}) = \sum_{i=1}^n \log f(x_i|\boldsymbol{\xi})$ and $B(\boldsymbol{\xi}_0, \delta)$ is a neighborhood of diameter δ centered at $\boldsymbol{\xi}_0$.

(C2) Let $\pi(\boldsymbol{\xi}) = E_{\boldsymbol{\xi}_0} J_0(X_0, \boldsymbol{\xi})$. The Jacobian function $J(\boldsymbol{X}, \boldsymbol{\xi}) \overset{a.s.}{\to} \pi(\boldsymbol{\xi})$ uniformly on compacts in $\boldsymbol{\xi}$. In the single variable case, this reduces to assumptions that $J(\boldsymbol{X}, \xi)$ is continuous in ξ, $\pi(\xi)$ is finite and $\pi(\xi_0) > 0$, and for some δ_0

$$E_{\xi_0}\left(\sup_{\xi \in B(\xi_0, \delta)} J_0(\boldsymbol{X}, \xi)\right) < \infty.$$

The extension to the multi-parameter case follows Yeo and Johnson [20] and replaces assumption C2 with C2.a,b, and c. Let $\omega \in \Omega$ be a collection of indices in $\{1, 2, \ldots p\}$ and $\bar{\omega} = \{1, 2, \ldots p\} \backslash \omega$. Define

$$J_\omega(\boldsymbol{x}_\omega; \boldsymbol{\xi}) = E_{\boldsymbol{\xi}_0}\left[J_0(\boldsymbol{x}_\omega, \boldsymbol{X}_{\bar{\omega}}; \boldsymbol{\xi})\right].$$

(C2.a) There exists an integrable and symmetric function $g(\cdot)$ and compact space $\bar{B}(\boldsymbol{\xi}_0, \delta)$ such that for $\boldsymbol{\xi} \in \bar{B}(\boldsymbol{\xi}_0, \delta)$ and $\boldsymbol{x} \in \mathbb{R}^p$ then $|J(\boldsymbol{x}; \boldsymbol{\xi})| \leq g(\boldsymbol{x})$.

(C2.b) There exists a sequence of measurable sets S_M^p such that

$$P(\mathbb{R}^p - \cup_{M=1}^\infty S_M^p) = 0.$$

(C2.c) For each M and for all $\omega \in \Omega$, $J_\omega(\boldsymbol{x}_\omega; \boldsymbol{\xi})$ is equicontinuous in $\boldsymbol{\xi}$ for $\{\boldsymbol{x}_\omega\} \in S_M^\omega$ where $S_M^p = S_M^\omega \times S_M^{\bar{\omega}}$.

Let $\mathcal{R}_{\boldsymbol{\xi}}$ be an observation from the fiducial distribution $r(\boldsymbol{\xi}|\boldsymbol{x})$ and denote the density of $s = \sqrt{n}\left(\mathcal{R}_{\boldsymbol{\xi}} - \hat{\boldsymbol{\xi}}_n\right)$ by $\pi^*(\boldsymbol{\xi}, \boldsymbol{x})$.

Theorem 1. *Given a random sample of independent observations $X_1, \ldots, X_n$, then under assumptions A0-A6, B1-B2, and C1-C2.c*

$$\int_{\mathbb{R}^p} \left| \pi^*(\boldsymbol{s}, \boldsymbol{x}) - \frac{\sqrt{det\,|I(\boldsymbol{\xi}_0)|}}{\sqrt{2\pi}} e^{-\boldsymbol{s}^T I(\boldsymbol{\xi}_0)\mathbf{s}/2} \right| d\boldsymbol{s} \overset{P_{\boldsymbol{\theta}_0}}{\to} 0. \tag{2}$$

Due to its technical nature, we relegate the proof to Appendix A, section 2.

3. Fiducial free-knot splines

We consider the fiducial free-knot spline solution for splines of degree $p \geq 4$. We first derive the fiducial distribution using a simple set of spline basis functions so

that the derivatives necessary derivatives can be calculated for the Jacobian. We then address the asymptotic behavior of the solution by applying Thm:1 to this solution. We next consider the practical issue of creating a proposal distribution for the MCMC simulation. Finally, we conduct a simulation study to compare the fiducial method to the Bayesian solution with reference prior $\propto \sigma^{-2}$ in four scenarios.

3.1. Deriving the fiducial free-knot spline

Suppose data $\{x_i, y_i\}$ for $i \in [1, \ldots, n]$ are generated from

$$y_i = g\left(x_i|\boldsymbol{\alpha}, \boldsymbol{t}\right) + \sigma\epsilon_i$$

where $\epsilon_i \overset{iid}{\sim} N\left(0,1\right)$ and $g\left(x|\boldsymbol{\alpha}, \boldsymbol{t}\right)$ is a degree $p \geq 4$ spline with κ knot points denoted $\boldsymbol{t}$ and $p+\kappa+1$ polynomial coefficients $\boldsymbol{\alpha}$. We assume that κ is known, but the knot locations $\boldsymbol{t}$ are unknown and are the primary target of investigation. The spline can be written using many different basis functions, but computational ease, we consider the piecewise truncated polynomial basis

$$g(x_i|\boldsymbol{\alpha}, \boldsymbol{t}) = \sum_{j=0}^{p} \alpha_j x_i^j + \sum_{k=1}^{\kappa} \alpha_{p+k} \left(x_i - t_k\right)_+^p$$

where

$$\left(u\right)_+ = \begin{cases} 0 & \text{if } u < 0 \\ u & \text{otherwise} \end{cases}$$

is the truncation operator and has higher precedence than the exponentiation. This representation makes it clear that the response function changes form at each knot point. The following derivation of the fiducial solution could, in principle, be done using more numerically stable basis functions, but the derivatives become more complicated. Our early work on this problem implemented a purely numerical solution using the b-spline basis, but the lack of closed form representation prevented showing that Thm:1 holds.

We derive the fiducial solution the to free-knot spline solution by first inverting the data generating equation and subsequently solving for ϵ_i. The Jacobian is then found by taking the derivative (with respect to the parameters of interest) of the inversion result.

Specifically, we denote the inverse by $\boldsymbol{G}_0^{-1}\left(y_i, \boldsymbol{\xi}\right)$ and let $\boldsymbol{\xi} = \left\{\boldsymbol{\alpha}, \boldsymbol{t}, \sigma^2\right\}^T$. We recognize that

$$\epsilon_i = \boldsymbol{G}_0^{-1}\left(y_i, \boldsymbol{\xi}\right) = \frac{1}{\sigma}\left(y_i - g(x_i|\boldsymbol{\theta})\right)$$

and therefore the partial derivatives with respect to the parameters is

$$\frac{\partial \boldsymbol{G}_0^{-1}\left(y_i, \boldsymbol{\xi}\right)}{\partial \boldsymbol{\alpha}} = -\frac{1}{\sigma}\left(1, x_i, \ldots, x_i^p, \left(x_i - t_1\right)_+^p, \ldots, \left(x_i - t_\kappa\right)_+^p\right)$$

$$\frac{\partial \boldsymbol{G}_0^{-1}(y_i, \boldsymbol{\xi})}{\partial \boldsymbol{t}} = \frac{p}{\sigma}\left(\alpha_{p+1}(x_i - t_1)_+^{p-1}, \ldots, \alpha_{p+\kappa}(x_i - t_\kappa)_+^{p-1}\right)$$

$$\frac{\partial \boldsymbol{G}_0^{-1}(y_i, \boldsymbol{\xi})}{\partial \sigma^2} = -\frac{1}{2\sigma^3}(y_i - g(x_i|\boldsymbol{\theta}))$$

$$\frac{\partial \boldsymbol{G}_0^{-1}(y_i, \boldsymbol{\xi})}{\partial y_i} = \frac{1}{\sigma}$$

where we define $0^0 = 0$ for notational convenience. Let $\boldsymbol{y}_0 = \{y_{(1)}, \ldots, y_{(l)}\}$ where $l = p + \kappa + 2$ be any selection of data points that satisfies the necessary invertability criteria. The Jacobian using these data points $\boldsymbol{y}_0$ is therefore

$$J_0(\boldsymbol{y}_0, \boldsymbol{\xi}) = \left|\frac{1}{\sigma^2} p^\kappa \det\left[\begin{array}{ccc} \boldsymbol{B}_{\boldsymbol{\alpha}} & \boldsymbol{B}_{\boldsymbol{t}} & \boldsymbol{B}_{\sigma^2} \end{array}\right]\right|$$

where

$$\boldsymbol{B}_{\boldsymbol{\alpha}} = \begin{bmatrix} 1 & x_{(1)} & \ldots & x_{(1)}^p & (x_{(1)} - t_1)_+^p & \ldots & (x_{(1)} - t_\kappa)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{(l)} & \ldots & x_{(l)}^p & (x_{(l)} - t_1)_+^p & \ldots & (x_{(l)} - t_\kappa)_+^p \end{bmatrix},$$

$$\boldsymbol{B}_{\boldsymbol{t}} = \begin{bmatrix} \alpha_{1+p+1}\left(x_{(1)} - t_1\right)_+^{p-1} & \ldots & \alpha_{1+p+\kappa}\left(x_{(1)} - t_\kappa\right)_+^{p-1} \\ \vdots & \ddots & \vdots \\ \alpha_{1+p+1}\left(x_{(l)} - t_1\right)_+^{p-1} & \ldots & \alpha_{1+p+\kappa}\left(x_{(l)} - t_\kappa\right)_+^{p-1} \end{bmatrix},$$

and

$$\boldsymbol{B}_{\sigma^2} = \begin{bmatrix} -\frac{1}{2}\left(y_{(1)} - g\left(x_{(1)}|\boldsymbol{\theta}\right)\right) \\ \vdots \\ -\frac{1}{2}\left(y_{(l)} - g\left(x_{(l)}|\boldsymbol{\theta}\right)\right) \end{bmatrix}.$$

Because $\boldsymbol{B}_{\sigma^2}$ contains a subtraction of a linear combination of columns of $\boldsymbol{B}_\alpha$ and $\boldsymbol{B}_{\boldsymbol{t}}$, the subtraction does not change the determinant and therefore

$$\left|\frac{1}{\sigma^2} p^\kappa \det\left[\begin{array}{ccc} \boldsymbol{B}_{\boldsymbol{\alpha}} & \boldsymbol{B}_{\boldsymbol{t}} & \boldsymbol{B}_{\sigma^2} \end{array}\right]\right| = \left|\frac{1}{2\sigma^2} p^\kappa \det\left[\begin{array}{ccc} \boldsymbol{B}_{\boldsymbol{\alpha}} & \boldsymbol{B}_{\boldsymbol{t}} & \tilde{\boldsymbol{B}}_{\sigma^2} \end{array}\right]\right|$$

where

$$\tilde{\boldsymbol{B}}_{\sigma^2} = \begin{bmatrix} y_{(1)} \\ \vdots \\ y_{(l)} \end{bmatrix}.$$

However the question of which sets of indices satisfy the one-to-one, and invertibility requirements is not obvious. A sufficient condition is that the set of indices includes at least two observations from each inter-knot region. Because we are primarily interested in cases where the number of observations is much larger than the number of knots, this condition is not onerous.

Theorem 2. *Given $g(x|\boldsymbol{\alpha},\boldsymbol{t})$, a free-knot spline of degree 4 or greater with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{t}$ with truncated polynomial basis functions and observations with x_i a randomly selected element on some contiguous interval $[a,b]$ of $\mathbb{R}$ and $y_i = g(x_i|\boldsymbol{\alpha},\boldsymbol{t}) + \sigma\epsilon_i$ where $\epsilon_i \overset{iid}{\sim} N(0,1)$, define $\boldsymbol{\xi} = \left(\boldsymbol{\alpha},\boldsymbol{t},\sigma^2\right)$. Let $\pi^*(\boldsymbol{\xi},\boldsymbol{y})$ be the fiducial distribution of $\mathcal{R}_{\boldsymbol{\xi}}$. Then*

$$\int_{\mathbb{R}^p} \left| \pi^*(\boldsymbol{s},\boldsymbol{y}) - \frac{\sqrt{det\left|I\left(\boldsymbol{\xi}_0\right)\right|}}{\sqrt{2\pi}} e^{-\boldsymbol{s}^T I(\boldsymbol{\xi}_0)\mathbf{s}/2} \right| d\boldsymbol{s} \overset{P_{\boldsymbol{\theta}_0}}{\to} 0$$

Proof. It suffices to show that the free-knot spline satisfies assumptions A0-A6, B1-B2, C1-C2.c. These are shown in Appendix B, which is available at the author's website. □

A shortcoming of this proof is the requirement that $p \geq 4$, while most many free-knot spline applications are concerned with degree $p = 1$ or 2 splines.

3.2. Numerical evaluation of the fiducial density

There are two substantial challenges to numerical evaluation of the fiducial density. The first is that the Jacobian does not simplify to a "nice" expression utilizing all of the data. We use the suggestion of Hannig [6] to use the mean of randomly selected Jacobians as an estimate of $J(\boldsymbol{x},\boldsymbol{\xi})$. The second challenge is that the scaling constant in the denominator of equation 1 is intractable and we only know the fiducial distribution up to a scaling constant. This is the same numerical challenge found in evaluating a Bayesian posterior distribution and we use Markov Chain Monte Carlo (MCMC) methods to select a random sample from the fiducial distribution. The key step of the MCMC is to produce good proposal values, which is often difficult when model parameters are highly correlated. Unfortunately our choice to use the analytically convenient truncated polynomial basis functions results in numerically inconvenient correlated parameters.

If the knot point locations were known, then the fiducial distribution of the $\boldsymbol{\alpha}$ and σ^2 terms is known and is same as the Bayesian posterior distribution with reference prior distribution $\propto \sigma^{-2}$. More formally, letting $\boldsymbol{X} = [\boldsymbol{B}_\alpha, \boldsymbol{B}_t]$ be the design matrix with fixed and known knot points, the fiducial distribution is $\boldsymbol{\alpha}|\sigma^2, y \sim N(\hat{\boldsymbol{\alpha}}, V_{\boldsymbol{\alpha}})$ where $\hat{\boldsymbol{\alpha}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$ and $V_{\boldsymbol{\alpha}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}$. Similarly the marginal distribution of $\sigma^2|\boldsymbol{y}$ is a scaled inverse-χ^2 distribution, $\sigma^2|\boldsymbol{y} \sim \text{Inv-}\chi^2\left(n-p-\kappa-1, s^2\right)$ where s^2 is the usual mean squared error term $s^2 = (\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\alpha}})^T(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\alpha}})/(n-p-\kappa-1)$. We denote the product of these distributions as the *fixed fiducial distribution.*

Unfortunately the fiducial distribution of σ^2 and $\boldsymbol{\alpha}$ conditioned on the knot point locations $\boldsymbol{t}$ is not the above fixed fiducial distribution because the Jacobian term cannot be factored into terms that contain only $\boldsymbol{t}$ parameters or only $\boldsymbol{\alpha}$ terms. However, the fixed fiducial distribution do provide useful proposal distribution in a MCMC estimation.

The procedure for creating a proposed value in the Markov chain is to take the current knot locations and perturb them by adding a small amount of noise. The proposed knots are $\boldsymbol{t}^* = \boldsymbol{t} + \boldsymbol{u}^*$ where $\boldsymbol{u}^* \sim MVN\left(\boldsymbol{0}, \sigma_k^2 \boldsymbol{I}_k\right)$, $\boldsymbol{I}_k$ is the identity matrix and σ_k^2 is the tuning parameter for the MCMC and reflects how much each knot point is "jittered". We then take these proposed knot points and consider them as known and use the aforementioned fixed fiducial distributions to produce proposed values for σ^2 and then $\boldsymbol{\alpha}$.

These three proposal distributions are multiplied to create the total proposal distribution $T\left(\boldsymbol{\xi}^* | \boldsymbol{\xi}\right)$. For the given proposed set of parameters, if the ratio

$$r = \frac{f\left(\boldsymbol{y} | \boldsymbol{\xi}^*\right)}{f\left(\boldsymbol{y} | \boldsymbol{\xi}\right)} \frac{T\left(\boldsymbol{\xi} | \boldsymbol{\xi}^*\right)}{T\left(\boldsymbol{\xi}^* | \boldsymbol{\xi}\right)}$$

is greater than a $Uniform(0,1)$ random deviate, we accept the proposed value as the next value in the Markov chain, otherwise the current vector of parameters is used.

The use of the fixed fiducial distribution is similar in spirit to the method of DiMatteo, Genovese and Kass [2] where they integrate out the $\boldsymbol{\alpha}$ and σ^2 parameters and consider only the distribution of the knot points $\boldsymbol{t}$. The difference is that their prior factored nicely whereas the Jacobian does not.

3.3. Simulation study for degree four splines

The simulation study will compare the fiducial method to the Bayesian method on four different degree four splines, all defined on domain $x \in [0,1]$ and with a similar range of y values.

The software we used to evaluate the performance of the fiducial solution compared to the Bayesian method with prior $\propto \sigma^{-2}$ used the same software for implementing the MCMC and generating proposed values, with the only difference in the software being whether the likelihood was multiplied by the Bayesian prior distribution or the calculated Jacobian.

The first spline has a single knot point at the center of the range of x values. The second has three knot points even spread through the x values. The third function also has three knot points, but the knots are not evenly distributed across the x values, instead they are clustered towards the left. The final function has three knot points evenly spread on the x-axis, but has a subtle change to the function at the first knot point, a larger change at the middle knot point and a large change at final knot. These functions are shown in figure 1.

For each scenario, we compared the methods using two different levels of variance and two samples sizes. The sample sizes $n = \{40, 100\}$ were chosen to reflect real world cases of data scarcity and moderate abundance. The two variance levels reflect an idealistically low level of variance ($\sigma = 0.1$) and a more realistic "signal-to-noise" level ($\sigma = 0.25$) commonly seen in the authors' applied work.

We consider coverage rates (figure 2) of the fiducial credible intervals of the true knot point values. In the 'coverage plots' presented, the x-axis denotes

Scenario	Knot Point(s)	Spline Coefficients
Single Knot	0.5	$0, 8, -60, 144, -108, 256$
3 Knots - Simple	$0.25, 0.50, 0.75$	$0, 30, -203, 386, -179, -276, 854, 270$
3 Knots - Clustered	$0.20, 0.40, 0.60$	$-1, 47, -397, 967, -640, -510, 2002, -1043$
3 Knots - Subtle	$0.25, 0.50, 0.75$	$0, -3, 2, 1, 1, 10, -100, 600$

TABLE 1
Coefficients defining the four different simulation scenarios.

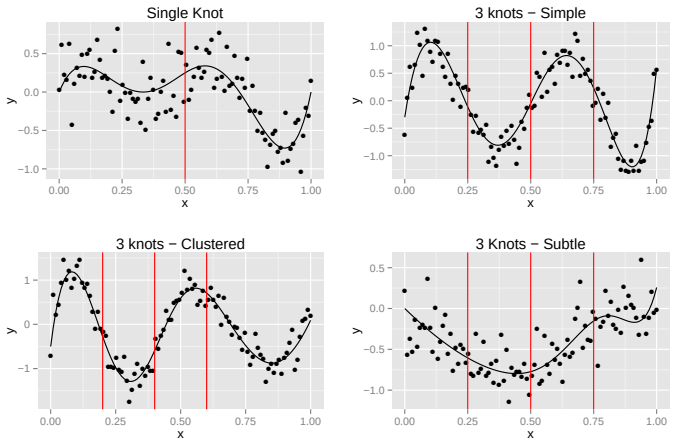

FIG 1. *Degree Four Examples - The examples shown are the high sample size and high variability case. Upper left panel - the "Single" knot case; upper right panel - 3 knots evenly spread across the x-axis which we refer to as the "simple" 3 knot case. Lower left panel - 3 knots "clustered" to the left side of the x-axis; lower right panel - 3 evenly spaced knots with with a "subtle" effect initially but with increasing effect size from left to right.*

the desired confidence level and the y-axis is the observed coverage rate in the experiment. If the observed coverage rate is below the equivalence line ($y = x$), then the method is considered *liberal* and if the observed rate is above the equivalence line then the method is *conservative*. Ideally a method would lie exactly on the equivalence line but a conservative method is more preferable to a liberal because claiming a 95% coverage rate when in truth the coverage rate is less is a more serious error than having the true coverage rate being larger than claimed. The only complaint against a conservative method is that the lengths of confidence intervals are larger than necessary to achieve the desired confidence level.

In the coverage plots presented, the oval lines around the equivalence line are the region in which we would expect the coverage rates to lie in due to stochastic variation in the simulation. For each simulation, the α-level necessary for the inclusion of the true parameter value in a confidence interval was calculated. Since the data is actually generated from the model we are fitting, then these α-levels should follow a uniform distribution if the coverage rates are correct. The jth ordered statistic of these therefore follows a $Beta\left(j, n-j+1\right)$ distribution and appropriate 95% point-wise confidence region can be calculated from this.

For each of the 16 combinations of function type, sample size and variance, 1000 simulations were performed and took approximately four days to run on a desktop computer. For the three knot simple case, a fiducial analysis took ≈ 100 seconds while the Bayesian solution took ≈ 10 seconds. The reason for this drastic difference is that for every evaluation of the fiducial density, the jacobian at that point must be estimated from averaging repeated samples of $J_0\left(\mathbf{x}_0, \xi\right)$.

3.4. Simulation Results

We display only the results of the "3 Knot - Clustered" function here and graphics of the other functions to the appendix because the results were similar.

The coverage rates (figure 2) for the for the fiducial method was typically slightly higher than the desired level, but was generally within the expected coverage region given the sample size. The Bayesian method also generally consistent with the desired rate, but was liberal in a few instances. For the "single knot case", both the fiducial and Bayesian methods were neither conservative nor liberal. In the "simple three knot case", the Bayesian method was liberal for all knots and sample sizes in the high variance cases, while the fiducial method was liberal for only the first knot in the high variance high sample case. In the "three knot clustered" case, the Bayesian method is conservative for knots one and two, but liberal for the third. In contrast the fiducial intervals were conservative for knot one. In the "three knot subtle" case the Bayesian method was conservative for knot one and two. The fiducial method was conservative for knot one in the small variance case. Overall, the fiducial estimator tends to have a coverage rate that is closer to the nominal rate than the Bayesian.

The lengths of the 95% confidence (or credible) interval lengths showed a consistent trend across our simulation (figure 3). The Bayesian intervals were

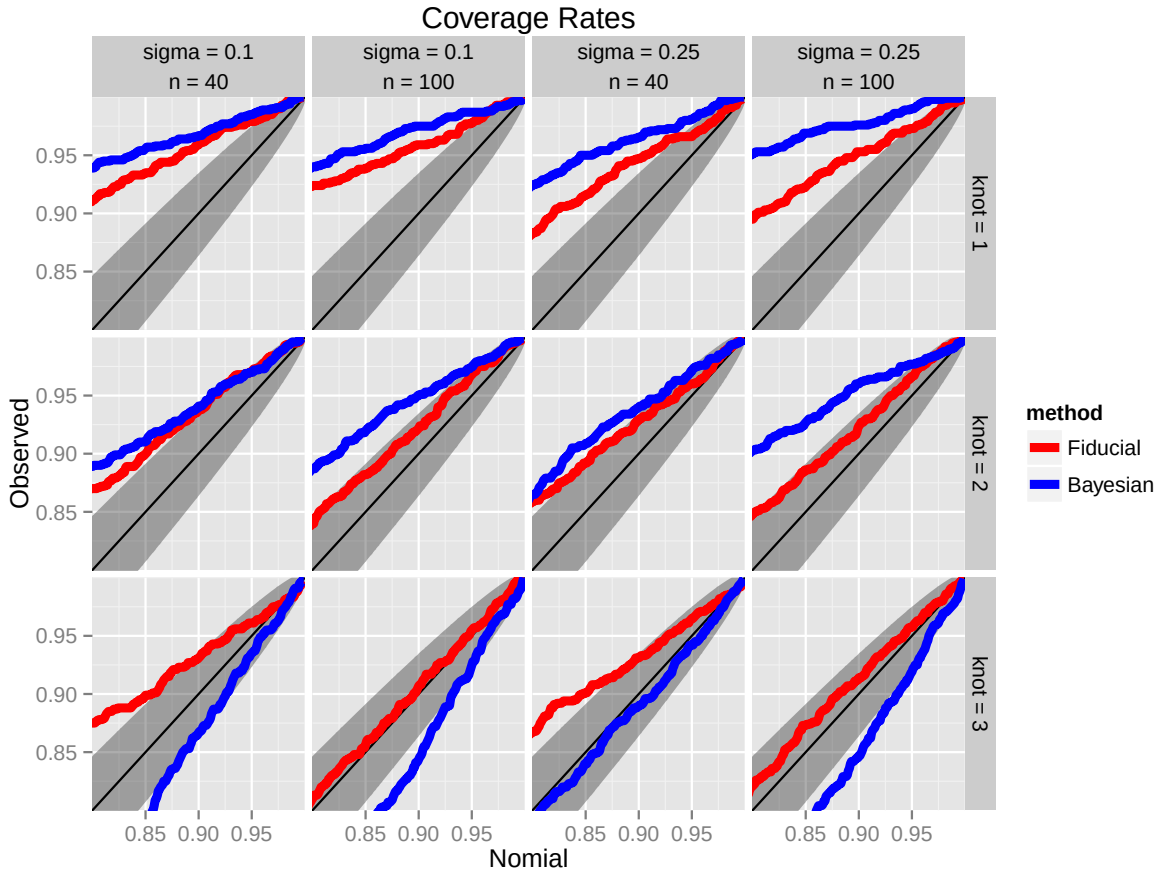

FIG 2. *Coverage rates for the "3 Knot - Clustered" simulation. The color (red, blue) represents the method (fiducial, Bayesian). The topmost panel is the coverage of knot 1 in the sigma* $= 0.1, n = 40$ *simulation. Graphs of the coverage rate for the other scenarios was similar and can be found in Appendix B.*

longer in every scenario we examined, however the difference was the smallest in the single knot case.

4. Conclusions

Free-knot splines are computationally challenging to fit, but in instances where inference on the knot points is desired, we believe that the fiducial method is a viable method for analysis. Simulation shows that the fiducial method is an effective method for the high degree free-knot spline problem and is superior to the Bayesian solution with prior $\propto \sigma^{-2}$. This is consistent with our previous experience of the fiducial method being equivalent to or better than the standard Bayesian solution derived using default prior [1].

The foundational theory for fiducial inference is given in [6] and this paper expands the fiducial Bernstein-von Mises theorem to the multivariate setting. However, this result is not the most general result possible due to the restrictive assumption of continuous fourth derivatives. In particular we believe that replacing the standard differentiability conditions used in the proof of Theorem 1 with Le Cam's continuity in quadratic mean assumptions [16] would allow us to relax the differentiability assumptions to obtain the most general Bernstein-von Mises type theorem for fiducial distributions. This is a subject of future work.

One case where continuous derivatives do not exist is the case of free-knot splines of degree one. These are of great interest due to the interpretability of the knot point as a change point. Based on our simulations results we conjecture that asymptotic normality holds even in this case. Further investigation into the the behavior of the fiducial method in this case relative to both the Bayesian solution and segmented regression [11] are of interest.

For this paper, we assume that the number of knot points to be fit is known. In some cases where the physical system under investigation provides insight into the number of knots. In the cases where the number of knots is not known, a reversible jump MCMC algorithm could allow for model selection, but would require some penalty term on models of increasing complexity.

Perhaps the largest reason for practitioners to not use new methodologies is the lack of accessible software packages. If a new methodology has no freely available software, or requires expensive software packages (such as Matlab and its associated toolboxes) applied researchers tend to not adopt a method. To alleviate this issue, we have provided the R package 'FiducialFreeKnotSplines' that contains the software used in the simulation studies conducted for this paper and is freely available on the Comprehensive R Archive Network (CRAN).

Acknowledgements

Dr Hannig thanks Prof. Hira Koul for his encouragement and help ever since he was a graduate student at Michigan State University. A young researcher cannot ask for a better role model. The authors also thank the two anonymous referees that made several useful suggestions for improving the manuscript.

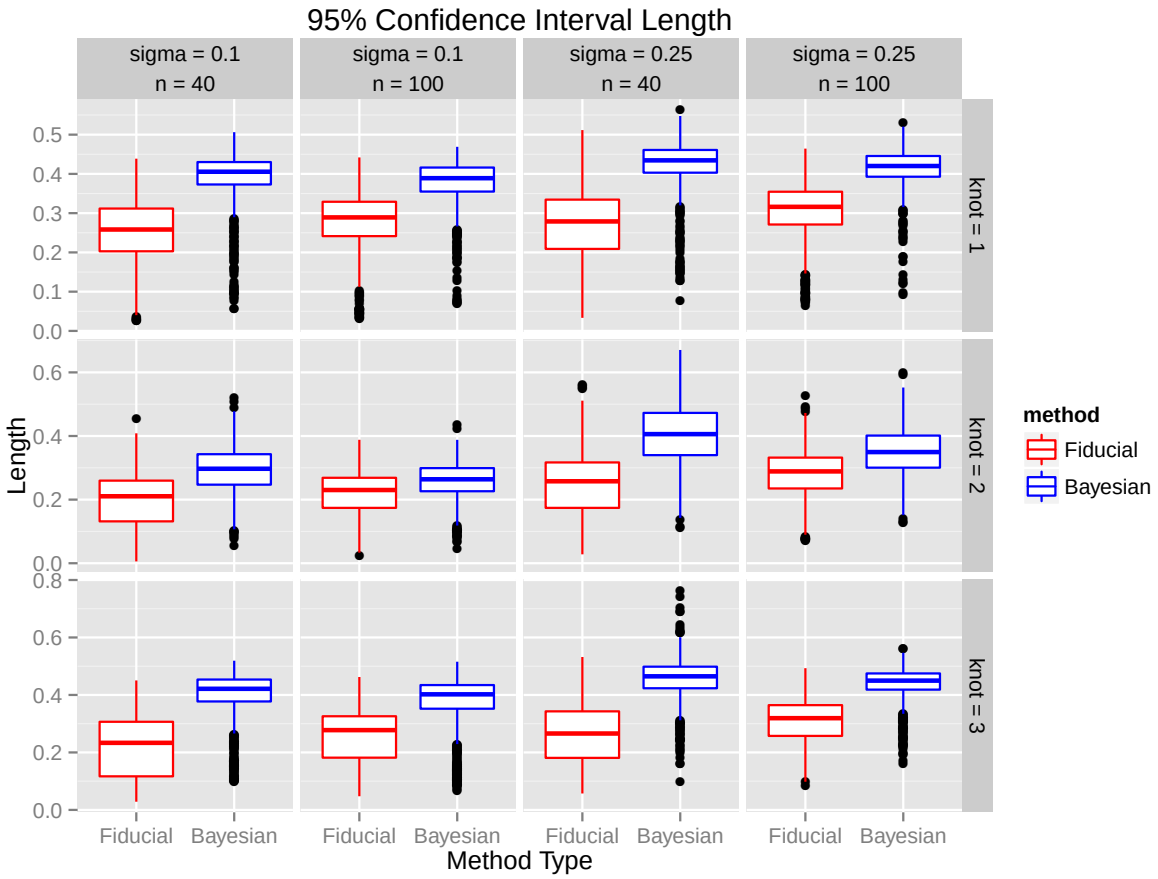

FIG 3. *Confidence interval lengths for the "3 Knot - Clustered" simulation. The color (red, blue) represents the method (fiducial, Bayesian). The topmost panel is the coverage of knot 1 in the sigma* $= 0.1, n = 40$ *simulation. Graphs of the interval lengths for the other scenarios was similar and can be found in Appendix B.*

References

[1] Cisewski, J. and Hannig, J. (2012). Generalized Fiducial Inference for Normal Linear Mixed Models. *The Annals of Statistics* **40** 2102–2127.

[2] DiMatteo, I., Genovese, C. R. and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88** 1055–1071.

[3] E, L., Hannig, J. and Iyer, H. (2008). Fiducial Intervals for Variance Components in an Un- balanced Two-component Normal Mixed Linear Model. *Journal of the American Statistical Association* **103** 854–865.

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

[5] Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian nonparametrics*. Spring-Verlag.

[6] Hannig, J. (2009). On generalized fiducial inference. *Statist Sinica* **19** 491–544.

[7] Hannig, J. (2013). Generalized Fiducial Inference via Discretization. *Statist Sinica* **23** 489–514.

[8] Hannig, J., Iyer, H. and Patterson, P. (2006). Fiducial Generalized Confidence Intervals. *Journal of the American Statistical Association* **101** 254–269.

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

[10] Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*. Springer, New York.

[11] Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine* **22** 3055–3071.

[12] Muggeo, V. M. R. (2008). Segmented: an R package to fit regression models with broken-line relationships. *R news* **8**.

[13] Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

[14] Sonderegger, D. L., Wang, H., Clements, W. H. and Noon, B. R. (2009). Using SiZer to detect thresholds in ecological data. *Frontiers in Ecology and the Environment* **7** 190–195.

[15] Toms, J. D. and Lesperance, M. L. (2003). Piecewise regression: A tool for identifying ecological thresholds. *Ecology* **84** 2034–2041.

[16] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

[17] Wandler, D. V. and Hannig, J. (2011). Generalized fiducial confidence intervals for extremes. *Extremes* **15** 67–87.

[18] Wandler, D. V. and Hannig, J. (2012). A fiducial approach to multiple comparisons. *Journal of Statistical Planning and Inference* **142** 878–895.

[19] Weerahandi, S. (1993). Generalized confidence intervals. *J. Amer. Statist. Assoc.* **88(423)** 899–905.

[20] Yeo, I. K. and Johnson, R. A. (2001). A uniform strong law of large numbers for U-statistics with application to transforming to near symmetry. *Statistics and Probability Letters* **51** 63–69.

Appendix A: Proof of asymptotic normality of fiducial estimators

We start with several assumptions. The assumptions A0-A6 are sufficient for the maximum likelihood estimate to converge asymptotically to a normal distribution and can be found in Lehmann and Casella [10] as 6.3 (A0)-(A2) and 6.5 (A)-(D). The assumption B2 shows that the Jacobian converges to a prior [6] and B1 is the assumption necessary for the Bayesian solution to converges to that of the MLE [5, Theorem 1.4.1].

A.1. Assumptions

A.1.1. Conditions for asymptotic normality of the MLE

(A0) The distributions $P_{\boldsymbol{\xi}}$ are distinct.

(A1) The set $\{x : f(x|\boldsymbol{\xi}) > 0\}$ is independent of the choice of $\boldsymbol{\xi}$.

(A2) The data $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ are iid with probability density $f(\cdot|\boldsymbol{\xi})$.

(A3) There exists an open neighborhood about the true parameter value $\boldsymbol{\xi}_0$ such that all third partial derivatives $\left(\partial^3/\partial\xi_i\partial\xi_j\partial\xi_k\right) f(\boldsymbol{x}|\boldsymbol{\xi})$ exist in the neighborhood, denoted by $B(\boldsymbol{\xi}_0, \delta)$.

(A4) The first and second derivatives of $L(\boldsymbol{\xi}, x) = \log f(x|\boldsymbol{\xi})$ satisfy

$$E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\xi_j}L(\boldsymbol{\xi}, x)\right] = 0$$

and

$$\begin{aligned} I_{j,k}(\boldsymbol{\xi}) &= E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\xi_j}L(\boldsymbol{\xi}, x)\cdot\frac{\partial}{\partial\xi_k}L(\boldsymbol{\xi}, x)\right] \\ &= -E_{\boldsymbol{\xi}}\left[\frac{\partial^2}{\partial\xi_j\partial\xi_k}L(\boldsymbol{\xi}, x)\right]. \end{aligned}$$

(A5) The information matrix $I(\boldsymbol{\xi})$ is positive definite for all $\boldsymbol{\xi} \in B(\boldsymbol{\xi}_0, \delta)$

(A6) There exists functions $M_{jkl}(\boldsymbol{x})$ such that

$$\sup_{\boldsymbol{\xi}\in B(\boldsymbol{\xi}_0,\delta)}\left|\frac{\partial^3}{\partial\xi_j\partial\xi_k\partial\xi_l}L(\boldsymbol{\xi}, x)\right| \le M_{j,k,l}(x) \quad \text{and} \quad E_{\boldsymbol{\xi}_0}M_{j,k,l}(x) < \infty$$

A.1.2. Conditions for the Bayesian posterior distribution to be close to that of the MLE.

Let $\pi(\boldsymbol{\xi}) = E_{\boldsymbol{\xi}_0}J_0(X_0, \boldsymbol{\xi})$ and $L_n(\boldsymbol{\xi}) = \sum L(\boldsymbol{\xi}, X_i)$

(B1) For any $\delta > 0$ there exists $\epsilon > 0$ such that

$$P_{\boldsymbol{\xi}_0}\left\{\sup_{\boldsymbol{\xi}\notin B(\boldsymbol{\xi}_0,\delta)}\frac{1}{n}\left(L_n(\boldsymbol{\xi}) - L_n(\boldsymbol{\xi}_0)\right) \le -\epsilon\right\} \to 1$$

(B2) $\pi(\boldsymbol{\xi})$ is positive at $\boldsymbol{\xi}_0$

A.1.3. Conditions for showing that the fiducial distribution is close to the Bayesian posterior

(C1) For any $\delta > 0$

$$\inf_{\boldsymbol{\xi} \notin B(\boldsymbol{\xi}_0, \delta)} \frac{\min_{i=1\ldots n} L(\boldsymbol{\xi}, X_i)}{|L_n(\boldsymbol{\xi}) - L_n(\boldsymbol{\xi}_0)|} \xrightarrow{P_{\boldsymbol{\xi}_0}} 0$$

(C2) Let $\pi(\boldsymbol{\xi}) = E_{\boldsymbol{\xi}_0} J_0(X_0, \boldsymbol{\xi})$. The Jacobian function $J(\boldsymbol{X}, \boldsymbol{\xi}) \overset{a.s.}{\to} \pi(\boldsymbol{\xi})$ uniformly on compacts in $\boldsymbol{\xi}$. In the single variable case, this reduces to $J(\boldsymbol{X}, \xi)$ is continuous in ξ, $\pi(\xi)$ is finite and $\pi(\xi_0) > 0$, and for some δ_0

$$E_{\xi_0}\left(\sup_{\xi \in B(\xi_0, \delta)} J_0(\boldsymbol{X}, \xi)\right) < \infty.$$

In the multivariate case, we follow Yeo and Johnson (2001). Let

$$J_j(x_1, \ldots, x_j; \boldsymbol{\xi}) = E_{\boldsymbol{\xi}_0}\left[J_0(x_1, \ldots, x_j, X_{j+1}, \ldots, X_k; \boldsymbol{\xi})\right].$$

(C2.a) There exists a integrable and symmetric functions $g(x_1, \ldots, x_j)$ and compact space $\bar{B}(\boldsymbol{\xi}_0, \delta)$ such that for $\boldsymbol{\xi} \in \bar{B}(\boldsymbol{\xi}_0, \delta)$ then $|J_j(x_1, \ldots, x_j; \boldsymbol{\xi})| \leq g(x_1, \ldots, x_j)$ for $j = 1, \ldots, k$.

(C2.b) There exists a sequence of measurable sets S_M^k such that

$$P\left(\mathbb{R}^k - \cup_{M=1}^{\infty} S_M^k\right) = 0$$

(C2.c) For each M and for all $j \in 1, \ldots, k$, $J_j(x_1, \ldots, x_j; \boldsymbol{\xi})$ is equicontinuous in $\boldsymbol{\xi}$ for $\{x_1, \ldots, x_j\} \in S_M^j$ where $S_M^k = S_M^j S_M^{k-j}$.

A.2. Proof of asymptotic normality of multivariate fiducial estimators

We now prove the the asymptotic normality (theorem 1) for multivariate Fiducial estimators.

Proof. Assume without loss of generality that $\xi \in \boldsymbol{\Xi} = \mathbb{R}^p$. We denote $J_n(\boldsymbol{x}_n, \boldsymbol{\xi})$ as the average of all possible Jacobians over a sample of size n and $\pi(\boldsymbol{\xi}) = E_{\boldsymbol{\xi}_0} J_0(\boldsymbol{x}, \boldsymbol{\xi})$. Assumption C2 and the uniform strong law of large numbers for U-statistics imply that $J_n(\boldsymbol{x}, \boldsymbol{\xi}) \overset{a.s.}{\to} \pi(\boldsymbol{\xi})$ uniformly in $\boldsymbol{\xi} \in \bar{B}(\boldsymbol{\xi}_0, \delta)$ and that $\pi(\boldsymbol{\xi})$ is continuous. Therefore

$$\sup_{\boldsymbol{\xi} \in \bar{B}(\boldsymbol{\xi}_0, \delta)} |J_n(x_n, \boldsymbol{\xi}) - \pi(\boldsymbol{\xi})| \to 0 \;\; P_{\boldsymbol{\xi}_0}\, a.s.$$

The multivariate proof now proceeds in a similar fashion as the univariate case. Let

$$\begin{aligned}
\pi^*(\boldsymbol{s},\boldsymbol{x}) &= \frac{J_n\left(\boldsymbol{x}_n,\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{s}}{\sqrt{n}}\right) f\left(\boldsymbol{x}_n|\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{s}}{\sqrt{n}}\right)}{\int_{\mathbb{R}^p} J_n\left(\boldsymbol{x}_n,\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{t}}{\sqrt{n}}\right) f\left(\boldsymbol{x}_n|\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{t}}{\sqrt{n}}\right) d\boldsymbol{t}} \\
&= \frac{J_n\left(\boldsymbol{x}_n,\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{s}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{s}}{\sqrt{n}}\right)\right]}{\int_{\mathbb{R}^p} J_n\left(\boldsymbol{x}_n,\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{t}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{t}}{\sqrt{n}}\right)\right] d\boldsymbol{t}} \\
&= \frac{J_n\left(\boldsymbol{x}_n,\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{s}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{s}}{\sqrt{n}}\right)-L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right]}{\int_{\mathbb{R}^p} J_n\left(\boldsymbol{x}_n,\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{t}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{t}}{\sqrt{n}}\right)-L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] d\boldsymbol{t}}
\end{aligned}$$

and just as Ghosh and Ramamoorthi [5], we let $H=-\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}}L_n\left(\hat{\boldsymbol{\xi}}_n\right)$ and we notice that $H\to I(\boldsymbol{\xi}_0)$ *a.s.* $P_{\boldsymbol{\xi}_0}$. It will be sufficient to prove

$$\begin{aligned}
\int_{\mathbb{R}^p}\Bigg|J_n\left(\boldsymbol{x}_n,\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{t}}{\sqrt{n}}\right)\exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n+\frac{\boldsymbol{t}}{\sqrt{n}}\right)-L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] \\
-\pi(\boldsymbol{\xi}_0)\exp\left[\frac{-\boldsymbol{t}^T I(\boldsymbol{\xi}_0)\,\boldsymbol{t}}{2}\right]\Bigg| d\boldsymbol{t} \quad \overset{P_{\boldsymbol{\xi}_0}}{\to} \quad 0 \qquad (3)
\end{aligned}$$

Let t_i represent the ith component of vector $\boldsymbol{t}$. By Taylor's Theorem, we can compute

$$\begin{aligned}
L_n\left(\hat{\boldsymbol{\xi}}_n+\boldsymbol{t}/\sqrt{n}\right) &= L_n\left(\hat{\boldsymbol{\xi}}_n\right)+\sum_{i=1}^{p}\left(\frac{t_i}{\sqrt{n}}\right)\frac{\partial}{\partial\xi_i}L_n\left(\hat{\boldsymbol{\xi}}_n\right) \\
&\quad+\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\left(\frac{t_it_j}{(\sqrt{n})^2}\frac{\partial}{\partial\xi_i\partial\xi_j}L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right) \\
&\quad+\frac{1}{6}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\left(\frac{t_it_jt_k}{(\sqrt{n})^3}\frac{\partial}{\partial\xi_i\partial\xi_j\partial\xi_k}L_n\left(\boldsymbol{\xi}'\right)\right) \\
&= L_n\left(\hat{\boldsymbol{\xi}}_n\right)-\frac{\boldsymbol{t}^TH\boldsymbol{t}}{2}+R_n
\end{aligned}$$

for some $\boldsymbol{\xi}'\in\left[\hat{\boldsymbol{\xi}}_n,\hat{\boldsymbol{\xi}}_n+\boldsymbol{t}/\sqrt{n}\right]$. Notice that $R_n=O_p\left(\|\boldsymbol{t}\|/n^{3/2}\right)$.

Given any $0<\delta<\delta_0$ and $c>0$, we break $\mathbb{R}^p$ into three regions:

$$\begin{aligned}
A_1 &= \left\{\boldsymbol{t}:\ \|\boldsymbol{t}\|<c\log\sqrt{n}\right\} \\
A_2 &= \left\{\boldsymbol{t}:\ c\log\sqrt{n}<\|\boldsymbol{t}\|<\delta\sqrt{n}\right\} \\
A_3 &= \left\{\boldsymbol{t}:\ \delta\sqrt{n}<\|\boldsymbol{t}\|\right\}
\end{aligned}$$

On $A_1 \cup A_2$ we compute

$$\begin{aligned}
&\int_{A_1\cup A_2} \left| J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] \right. \\
&\qquad \left. -\pi\left(\boldsymbol{\xi}_0\right) \exp\left[-\frac{1}{2}\boldsymbol{t}' I\left(\boldsymbol{\xi}_0\right)\boldsymbol{t}\right] \right| d\boldsymbol{t} \\
\leq\ &\int_{A_1\cup A_2} \left| J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) - \pi\left(\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) \right| \\
&\qquad \cdot \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] d\boldsymbol{t} \\
&+ \int_{A_1\cup A_2} \left| \pi\left(\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] \right. \\
&\qquad \left. -\pi\left(\boldsymbol{\xi}_0\right) \exp\left[-\frac{1}{2}\boldsymbol{t}' I\left(\boldsymbol{\xi}_0\right)\boldsymbol{t}\right] \right| d\boldsymbol{t}
\end{aligned}$$

Since $\pi(\cdot)$ is a proper prior on $A_1 \cup A_2$, then the second term goes to 0 by the Bayesian Bernstein-von Mises theorem. Next we notice that

$$\begin{aligned}
&\int_{A_1\cup A_2} \left| J_n\left(x, \hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) - \pi\left(\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) \right| \\
&\qquad \cdot \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] d\boldsymbol{t} \\
\leq\ &\sup_{\boldsymbol{t}\in A_1\cup A_2} \left| J_n\left(x, \hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) - \pi\left(\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) \right| \\
&\qquad \cdot \int_{A_1\cup A_2} \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] d\boldsymbol{t}
\end{aligned}$$

Since $\sqrt{n}\left(\hat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0\right) \xrightarrow{\mathcal{D}} N\left(0, I\left(\boldsymbol{\xi}_0\right)^{-1}\right)$, then

$$P_{\boldsymbol{\xi}_0}\left[\left\{\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n};\ \boldsymbol{t} \in A_1 \cup A_2\right\} \subset B\left(\boldsymbol{\xi}_0, \delta_0\right)\right] \to 1.$$

Furthermore, since $L_n\left(\hat{\boldsymbol{\xi}}_n + \boldsymbol{t}/\sqrt{n}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right) = -\frac{\boldsymbol{t}^T H \boldsymbol{t}}{2} + R_n$ then the integral converges in probability to 1. Since $\max_{\boldsymbol{t}\in A_1\cup A_2} \|\boldsymbol{t}/\sqrt{n}\| \leq \delta$ and $J_n \to \pi$, then the term $\to 0$ in probability.

Next we turn to

$$\begin{aligned}
&\int_{A_3} \left| J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] \right. \\
&\qquad \left. -\pi\left(\boldsymbol{\xi}_0\right) \exp\left[\frac{-\boldsymbol{t}^T I\left(\boldsymbol{\xi}_0\right)\boldsymbol{t}}{2}\right] \right| d\boldsymbol{t} \\
\leq\ &\int_{A_3} J_n\left(\boldsymbol{x}_i, \hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] d\boldsymbol{t} \\
&+ \int_{A_3} \pi\left(\boldsymbol{\xi}_0\right) \exp\left[\frac{-\boldsymbol{t}^T I\left(\boldsymbol{\xi}_0\right)\boldsymbol{t}}{2}\right] d\boldsymbol{t}
\end{aligned}$$

The second integral goes to 0 in $P_{\boldsymbol{\xi}_0}$ probability because $\min_{A_3} \|\boldsymbol{t}\| \to \infty$. As for the first integral,

$$\begin{aligned}
&\int_{A_3} J_n\left(\boldsymbol{x}, \hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] d\boldsymbol{t} \\
&= \frac{1}{n}\sum_{i=1}^{n} \int_{A_3} J\left(\boldsymbol{x}_i, \hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] d\boldsymbol{t} \\
&= \frac{1}{n}\sum_{i=1}^{n} \int_{A_3} J\left(\boldsymbol{x}_i, \hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) f\left(x_i | \hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) \\
&\qquad \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right) - \log f\left(x_i | \hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right)\right] d\boldsymbol{t}
\end{aligned}$$

Because $J(\cdot)$ is a probability measure, then so is $J(\cdot) f(\cdot)$. Assumption C1 assures that the exponent goes to $-\infty$ and therefore the integral converges to 0 in probability.

Having shown Eq:3, we now follow Ghosh and Ramamoorthi [5] and let

$$C_n = \int_{\mathbb{R}^p} \left| J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{t}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{t}}{\sqrt{n}}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] \right| d\boldsymbol{t}$$

then the main result to be proved (Eq:2) becomes

$$\begin{aligned}
C_n^{-1} \Bigg\{ \int_{\mathbb{R}^p} \Bigg| J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] \cdot \\
- C_n \frac{\sqrt{det\left|I\left(\boldsymbol{\xi}_0\right)\right|}}{\sqrt{2\pi}} e^{-\boldsymbol{s}^T I(\boldsymbol{\xi}_0)\boldsymbol{s}/2} \Bigg| \Bigg\} d\boldsymbol{s} \overset{P_{\boldsymbol{\xi}_0}}{\to} 0 \quad (4)
\end{aligned}$$

Because

$$\begin{aligned}
\int_{\mathbb{R}^p} J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n\right) \exp\left[-\frac{\boldsymbol{s}^T H \boldsymbol{s}}{2}\right] d\boldsymbol{s} &= J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n\right) \int_{\mathbb{R}^p} \exp\left[-\frac{\boldsymbol{s}^T H \boldsymbol{s}}{2}\right] d\boldsymbol{s} \\
&= J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n\right) \frac{\sqrt{2\pi}}{\sqrt{\det(H)}} \\
&\overset{a.s.}{\to} \pi\left(\boldsymbol{\xi}_0\right) \sqrt{\frac{2\pi}{\det\left(I\left(\boldsymbol{\xi}_0\right)\right)}}
\end{aligned}$$

and Eq:3 imply that $C_n \overset{P}{\to} \pi\left(\boldsymbol{\xi}_0\right) \sqrt{\frac{2\pi}{\det(I(\boldsymbol{\xi}_0))}}$ it is enough to show that the integral in Eq:4 goes to 0 in probability. This integral is less than $I_1 + I_2$ where

$$\begin{aligned}
I_1 = \int_{\mathbb{R}^P} \Bigg| &J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) \exp\left[L_n\left(\hat{\boldsymbol{\xi}}_n + \frac{\boldsymbol{s}}{\sqrt{n}}\right) - L_n\left(\hat{\boldsymbol{\xi}}_n\right)\right] \\
&- J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n\right) \exp\left[\frac{-\boldsymbol{s}^T H \boldsymbol{s}}{2}\right] \Bigg| d\boldsymbol{s}
\end{aligned}$$

and

$$I_2 = \int_{\mathbb{R}^P} \left| J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n\right) \exp\left[\frac{-\boldsymbol{s}^T H \boldsymbol{s}}{2}\right] - C_n \frac{\sqrt{det\left|I\left(\boldsymbol{\xi}_0\right)\right|}}{\sqrt{2\pi}} e^{-\boldsymbol{s}^T I(\boldsymbol{\xi}_0)\mathbf{s}/2} \right| d\boldsymbol{s}.$$

Eq:3 shows that $I_1 \to 0$ in probability and I_2 is

$$\begin{aligned} I_2 &= \left| J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n\right) - C_n \frac{\sqrt{det\left|I\left(\boldsymbol{\xi}_0\right)\right|}}{\sqrt{2\pi}} \right| \int_{\mathbb{R}^P} \exp\left[\frac{-\boldsymbol{s}^T H \boldsymbol{s}}{2}\right] d\boldsymbol{s} \\ &\xrightarrow{P} 0 \end{aligned}$$

because $J_n\left(\boldsymbol{x}_n, \hat{\boldsymbol{\xi}}_n\right) \xrightarrow{P} \pi\left(\boldsymbol{\xi}_0\right)$ and $C_n \xrightarrow{P} \pi\left(\boldsymbol{\xi}_0\right) \sqrt{\frac{2\pi}{\det(I(\boldsymbol{\xi}_0))}}$. □