Fiducial theory for free-knot splines

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

This is an on-line companion to the paper Fiducial theory for free-knot splines. It contains proof of Theorem 2 – verification of the conditions of Bernstein-von Mises theorem for multivariate normal distributions (Theorem 1) for free-knot splines of order 4 or higher. Additionally this on-line appendix contains more detailed simulation results.

Appendix B: Proof of assumptions for free-knot splines using a truncated polynomial basis.

We now consider the free-knot spline case. Suppose we are interested in a p degree (order $m = p+1$) polynomial spline with κ knot points, $\boldsymbol{t} = \{t_1, \ldots, t_\kappa\}^T$ where $t_k \in (a+\delta, b-\delta)$ and $|t_i - t_j| \leq \delta$ for $i \neq j$ and some $\delta > 0$. Furthermore, we assume that the data points $\{x_i, y_i\}$ independent with the distribution of the x_i having positive density on $[a, b]$.

Denote the truncated polynomial spline basis functions as

$$\begin{aligned} N(x,\boldsymbol{t}) &= \{N_1(x,\boldsymbol{t}), \ldots, N_{\kappa+m}(x,\boldsymbol{t})\}^T \\ &= \left\{1, x, \ldots, x^p, (x-t_1)_+^p, \ldots, (x-t_\kappa)_+^p\right\}^T \end{aligned}$$

and let $y_i = N(x_i, \boldsymbol{t})^T \boldsymbol{\alpha} + \sigma \epsilon_i$ where $\epsilon_i \overset{iid}{\sim} N(0,1)$ and thus the density function is

$$f(y, \boldsymbol{\xi}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(y - N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2\right]$$

where $\boldsymbol{\xi} = \{\boldsymbol{t}, \boldsymbol{\alpha}, \sigma^2\}$ and the log-likelihood function is

$$L(\boldsymbol{\xi}, y) = \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left(y - N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2$$

B.1. Assumptions A0-A4

Assumptions A0-A2 are satisfied. We now consider assumption A3 and A4. We note that if $p \geq 4$ then the necessary three continuous derivatives exist and now

[*]Corresponding Author: derek.sonderegger@nau.edu

[†]Department of Mathematics and Statistics, Northern Arizona Univeriesty

[‡]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.

examine the derivatives. Let $\boldsymbol{\theta} = \{\boldsymbol{t}, \boldsymbol{\alpha}\}$ and thus

$$\begin{aligned}
E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\theta_j}L(\boldsymbol{\xi},y)\right] &= E_{\boldsymbol{\xi}}\left[-\frac{1}{2\sigma^2}2\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\left(-\frac{\partial}{\partial\theta_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\right] \\
&= -\frac{1}{2\sigma^2}2\left(E_{\boldsymbol{\xi}}\left[y\right]-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\left(-\frac{\partial}{\partial\theta_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right) \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\sigma^2}L(\boldsymbol{\xi},\boldsymbol{y})\right] &= E_{\boldsymbol{\xi}}\left[-\frac{1}{2\sigma^2}+\frac{1}{2\left(\sigma^2\right)^2}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2\right] \\
&= -\frac{1}{2\sigma^2}+\frac{1}{2\left(\sigma^2\right)^2}\left(\sigma^2\right) \\
&= 0.
\end{aligned}$$

Next we consider information matrix. First we consider the $\boldsymbol{\theta}$ terms.

$$\begin{aligned}
&E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\theta_j}L(\boldsymbol{\xi},\boldsymbol{y})\,\frac{\partial}{\partial\theta_k}L(\boldsymbol{\xi},\boldsymbol{y})\right] \\
&= E_{\boldsymbol{\xi}}\left[\frac{1}{\sigma^4}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2\left(\frac{\partial}{\partial\theta_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\left(\frac{\partial}{\partial\theta_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\right] \\
&= \frac{1}{\sigma^4}E_{\boldsymbol{\xi}}\left[\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2\right]\left(\frac{\partial}{\partial\theta_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\left(\frac{\partial}{\partial\theta_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right) \\
&= \frac{1}{\sigma^2}\left(\frac{\partial}{\partial\theta_j}N(x_i,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\left(\frac{\partial}{\partial\theta_k}N(x_i,\boldsymbol{t})^T\boldsymbol{\alpha}\right)
\end{aligned}$$

The j,k partials for the second derivative are

$$\begin{aligned}
&\frac{\partial^2}{\partial\theta_j\partial\theta_k}L(\boldsymbol{\xi},y) \\
&= \frac{\partial}{\partial\theta_j}\left[-\frac{1}{2\sigma^2}2\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\left(-\frac{\partial}{\partial\theta_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\right] \\
&= \frac{\partial}{\partial\theta_j}\left[-\frac{1}{\sigma^2}\left(-y_i\left(\frac{\partial}{\partial\theta_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)+N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\left(\frac{\partial}{\partial\theta_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\right)\right] \\
&= -\frac{1}{\sigma^2}\left[-y\frac{\partial^2}{\partial\theta_j\partial\theta_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}+\left(\frac{\partial}{\partial\theta_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\left(\frac{\partial}{\partial\theta_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\right. \\
&\qquad\left.+N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\frac{\partial^2}{\partial\theta_j\partial\theta_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right]
\end{aligned}$$

which have expectation

$$\begin{aligned}
E_{\boldsymbol{\xi}}\left[\frac{\partial^2}{\partial\theta_j\partial\theta_k}L(\boldsymbol{\xi},y)\right] &= -\frac{1}{\sigma^2}\left(\frac{\partial}{\partial\theta_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\left(\frac{\partial}{\partial\theta_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right) \\
&= -E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\theta_j}L(\boldsymbol{\xi},y)\,\frac{\partial}{\partial\theta_k}L(\boldsymbol{\xi},\boldsymbol{y})\right]
\end{aligned}$$

as necessary. Next we consider

$$\begin{aligned}
&E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\theta_j}L(\boldsymbol{\xi},y)\,\frac{\partial}{\partial\sigma^2}L(\boldsymbol{\xi},y)\right]\\
&= E_{\boldsymbol{\xi}}\left[\frac{1}{\sigma^2}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\frac{\partial}{\partial\theta_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\left[-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2\right]\right]\\
&= E_{\boldsymbol{\xi}}\left[-\frac{1}{2\sigma^4}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\frac{\partial}{\partial\theta_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}+\frac{1}{2\sigma^6}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^3\frac{\partial}{\partial\theta_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right]\\
&= 0
\end{aligned}$$

which is equal to

$$\begin{aligned}
E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\theta_j\partial\sigma^2}L(\boldsymbol{\xi},\boldsymbol{y})\right] &= E_{\boldsymbol{\xi}}\left[\frac{2}{2\sigma^4}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\frac{\partial}{\partial\theta_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right]\\
&= 0.
\end{aligned}$$

Finally

$$\begin{aligned}
&E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\sigma^2}L(\boldsymbol{\xi},y)\,\frac{\partial}{\partial\sigma^2}L(\boldsymbol{\xi},y)\right]\\
&= E_{\boldsymbol{\xi}}\left[\left\{-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2\right\}\left\{-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2\right\}\right]\\
&= E_{\boldsymbol{\xi}}\left[\frac{1}{4\sigma^4}-\frac{2}{4\sigma^6}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2+\frac{1}{4\sigma^8}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^4\right]\\
&= \frac{1}{4\sigma_0^4}-\frac{2}{4\sigma_0^6}\sigma_0^2+\frac{1}{4\sigma_0^8}3\sigma_0^4\\
&= \frac{2}{4\sigma_0^4}
\end{aligned}$$

which is equal to

$$\begin{aligned}
-E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\sigma^2\partial\sigma^2}L(\boldsymbol{\xi},y)\right] &= -E_{\boldsymbol{\xi}}\left[\frac{1}{2}\sigma^{-4}-\frac{2}{2}\sigma^{-6}\left(y-N(x,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2\right]\\
&= -\frac{1}{2}\sigma_0^{-4}+\frac{2}{2}\sigma_0^{-4}.
\end{aligned}$$

Therefore the interchange of integration and differentiation is justified.

B.2. Assumptions A5

To address whether the information matrix is positive definite, we notice that since $E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\sigma^2}L(\boldsymbol{\xi},y)\,\frac{\partial}{\partial\sigma^2}L(\boldsymbol{\xi},y)\right] > 0$ and $E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\theta_j}L(\boldsymbol{\xi},y)\,\frac{\partial}{\partial\sigma^2}L(\boldsymbol{\xi},y)\right] = 0$, we only need to be concerned with the sub-matrix

$$\begin{aligned} I_{j,k}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} E_{\boldsymbol{\xi}}\left[\frac{\partial}{\partial\theta_j}L(\boldsymbol{\xi},y_i)\,\frac{\partial}{\partial\theta_k}L(\boldsymbol{\xi},y_i)\right] \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta_j}N(x_i,\boldsymbol{t})^T\boldsymbol{\alpha}\right)\left(\frac{\partial}{\partial\theta_k}N(x_i,\boldsymbol{t})^T\boldsymbol{\alpha}\right). \end{aligned}$$

where the σ^{-2} term can be ignored because it doesn't affect the positive definiteness. First we note

$$\frac{\partial}{\partial t_j}N(x_i,\boldsymbol{t})^T\boldsymbol{\alpha} = -p\,(x_i-t_j)_+^{p-1}\,\alpha_{p+j+1}$$

$$\frac{\partial}{\partial \alpha_j}N(x_i,\boldsymbol{t})^T\boldsymbol{\alpha} = N_j(x_i,\boldsymbol{t}).$$

If we let

$$X = \begin{bmatrix} N_1(x_1,\boldsymbol{t}) & \cdots & N_{m+\kappa}(x_1,\boldsymbol{t}) & \frac{\partial}{\partial t_1}N(x_1,\boldsymbol{t})^T\boldsymbol{\alpha} & \cdots & \frac{\partial}{\partial t_\kappa}N(x_1,\boldsymbol{t})^T\boldsymbol{\alpha} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ N_1(x_n,\boldsymbol{t}) & \cdots & N_{m+\kappa}(x_n,\boldsymbol{t}) & \frac{\partial}{\partial t_1}N(x_n,\boldsymbol{t})^T\boldsymbol{\alpha} & \cdots & \frac{\partial}{\partial t_\kappa}N(x_n,\boldsymbol{t})^T\boldsymbol{\alpha} \end{bmatrix}$$

then $I(\boldsymbol{\theta}) = X^T X$. Then $I(\boldsymbol{\theta})$ is positive definite if the columns of X are linearly independent. This is true under the assumptions that $t_j \neq t_k$ and that $\alpha_{m+j} \neq 0$.

B.3. Assumptions A6

We next consider a bound on the third partial derivatives. We start with the derivatives of the basis functions.

$$\frac{\partial^2}{\partial t_j \partial t_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha} = 0 \qquad \text{if } \ j \neq k$$

$$\frac{\partial^2}{\partial t_j \partial t_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha} = p(p-1)\,(x-t_j)_+^{p-2}\,\alpha_{p+j+1}$$

$$\frac{\partial^2}{\partial \alpha_j \partial \alpha_k}N(x,\boldsymbol{t})^T\boldsymbol{\alpha} = 0$$

$$\frac{\partial^2}{\partial t_j \partial \alpha_{p+j+1}}N(x,\boldsymbol{t})^T\boldsymbol{\alpha} = -p\,(x-t_j)_+^{p-1}$$

$$\frac{\partial^3}{\partial t_j \partial t_j \partial t_j}N(x,\boldsymbol{t})^T\boldsymbol{\alpha} = -p(p-1)(p-2)\,(x-t_j)_+^{p-3}\,\alpha_{p+j+1}$$

$$\frac{\partial^3}{\partial t_j \partial t_j \partial \alpha_{p+j+1}}N(x,\boldsymbol{t})^T\boldsymbol{\alpha} = p(p-1)\,(x-t_j)_+^{p-2}$$

Since x is an element of a compact set, then for $\boldsymbol{\xi} \in B(\boldsymbol{\xi}_0, \delta)$ all of the above partials are bounded as is $N(x, \boldsymbol{t})^T \boldsymbol{\alpha}$. Therefore

$$\begin{aligned}
&\frac{\partial^3}{\partial\theta_j \partial\theta_k \partial\theta_l} L(\boldsymbol{\xi}, x) \\
&= -\frac{1}{\sigma^2} \left[-y \frac{\partial^3}{\partial\theta_j \partial\theta_k \partial\theta_l} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} + \left(\frac{\partial^2}{\partial\theta_j \partial\theta_k} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right) \left(\frac{\partial^2}{\partial\theta_l} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right) \right. \\
&\quad + \left(\frac{\partial^2}{\partial\theta_j \partial\theta_l} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right) \left(\frac{\partial^2}{\partial\theta_k} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right) \\
&\quad + \left(\frac{\partial^2}{\partial\theta_l \partial\theta_k} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right) \left(\frac{\partial^2}{\partial\theta_j} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right) \\
&\quad \left. + N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \left(\frac{\partial^3}{\partial\theta_j \partial\theta_k \partial\theta_l} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right) \right]
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\partial^3}{\partial\theta_j \partial\theta_k \partial\sigma^2} L(\boldsymbol{\xi}, x) \\
&= \frac{1}{\sigma^4} \left[-y \frac{\partial^2}{\partial\theta_j \partial\theta_k} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} + \left(\frac{\partial}{\partial\theta_k} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right) \left(\frac{\partial}{\partial\theta_j} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right) \right. \\
&\qquad \left. + N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \frac{\partial^2}{\partial\theta_j \partial\theta_k} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right]
\end{aligned}$$

and

$$\frac{\partial^3}{\partial\theta_j \partial\sigma^2 \partial\sigma^2} L(\boldsymbol{\xi}, y) = -\frac{2}{\sigma^6} \left(y - N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right) \left(-\frac{\partial}{\partial\theta_j} N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right)$$

and

$$\frac{\partial^3}{\partial\sigma^2 \partial\sigma^2 \partial\sigma^2} L(\boldsymbol{\xi}, \boldsymbol{y}) = -\frac{1}{\sigma^6} + \frac{3}{\sigma^8} \left(y - N(x, \boldsymbol{t})^T \boldsymbol{\alpha} \right)^2$$

are also bounded $\boldsymbol{\xi} \in B(\boldsymbol{\xi}_0, \delta)$ since $\sigma_0^2 > 0$ by assumption. The expectation of the bounds also clearly exists.

B.4. Lemmas

To show that the remaining assumptions are satisfied, we first examine the behavior of

$$g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i) = N(x_i, \boldsymbol{t}_0)^T \boldsymbol{\alpha}_0 - N(x_i, \boldsymbol{t})^T \boldsymbol{\alpha}.$$

Notice that for x_i chosen on a uniform grid over $[a, b]$ then

$$\frac{1}{n} \sum_{i=1}^{n} \left(g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i) \right)^2 \quad \to \quad \frac{1}{b-a} \int_a^b \left(g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x) \right)^2 \, dx.$$

Furthermore we notice that $g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x)$ is also a spline. The sum of the two splines is also a spline. Consider the degree p case of $g(x|\boldsymbol{\alpha}, t) + g(x|\boldsymbol{\alpha}^*, t^*)$ where $t < t^*$. Then the sum is a spline with knot points $\{t, t^*\}$ and whose first $p+1$ coefficients are $\boldsymbol{\alpha} + \boldsymbol{\alpha}^*$ and last two coefficients are $\left\{\alpha_{p+1}, \alpha^*_{p+1}\right\}$.

At this point we also notice

$$\begin{aligned} E\left[n^{-1}\sum g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i)\,\epsilon_i\right] &= n^{-1}\sum g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i)\, E[\epsilon_i] \\ &= 0 \end{aligned}$$

$$\begin{aligned} V\left[n^{-1}\sum g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i)\,\epsilon_i\right] &= n^{-2} V\left[\sum g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i)\,\epsilon_i\right] \\ &= n^{-2}\sum V\left[g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i)\,\epsilon_i\right] \\ &= n^{-2}\sum g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i)^2\, V[\epsilon_i] \\ &= n^{-2}\sum g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i)^2 \\ &\to 0 \end{aligned}$$

and that $\sum \epsilon_i^2 \sim \chi^2_n$ and thus $n^{-1}\sum \epsilon_i^2$ converges in probability to the constant 1. Therefore, by the SLLN,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\left[g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i) + \sigma_0\epsilon_i\right]^2 &= \frac{1}{n}\sum_{i=1}^{n}\left[g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i)\right]^2 + \frac{2\sigma_0}{n}\sum_{i=1}^{n}\epsilon_i g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i) + \frac{\sigma_0^2}{n}\sum_{i=1}^{n}\epsilon_i^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i)\right]^2 + O_p\left(n^{-1}\right) + \frac{\sigma_0^2}{n}\sum_{i=1}^{n}\epsilon_i^2 \\ &\overset{a.s.}{\to} \frac{1}{b-a}\int_a^b \left(g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x)\right)^2\, dx + \sigma_0^2. \end{aligned}$$

Lemma 1. *Given a degree p polynomial $g(x|\boldsymbol{\alpha})$ on $[a,b]$ with coefficients $\boldsymbol{\alpha}$, then $\exists\, \lambda_{n,m}, \lambda_{n,M} > 0$ such that $||\boldsymbol{\alpha}||^2\lambda^2_{n,m} \le \frac{1}{n}\sum_{i=1}^{n}\left[g(x_i|\boldsymbol{\alpha})\right]^2 \le ||\boldsymbol{\alpha}||^2\lambda^2_{n,M}$.*

Proof. If $\boldsymbol{\alpha} = \mathbf{0}$, then $g(x|\boldsymbol{\alpha}) = 0$ and the result is obvious. If $g(x|\boldsymbol{\alpha})$ is a polynomial with at least one non-zero coefficient, it therefore cannot be identically zero on $[a,b]$ and therefore for $n > p$ then $\frac{1}{n}\sum\left[g(x_i|\boldsymbol{\alpha})\right]^2 > 0$ since the polynomial can only have at most p zeros. We notice that

$$\begin{aligned} \int_a^b \left[g(x|\boldsymbol{\alpha})\right]^2\, dx &= \int_a^b \left[\sum_{i=0}^{p}\alpha_i^2 x^{2i} + 2\sum_{i=0}^{p-1}\sum_{j=i+1}^{p}\alpha_i\alpha_j x^{i+j}\right] dx \\ &= \left.\sum_{i=0}^{p}\frac{\alpha_i^2}{i+1}x^{2i+1} + 2\sum_{i=0}^{p-1}\sum_{j=i+1}^{p}\frac{\alpha_i\alpha_j}{i+j+1}x^{i+j+1}\right|_{x=a}^{b} \\ &= \boldsymbol{\alpha}^T X \boldsymbol{\alpha} \end{aligned}$$

where the matrix $\boldsymbol{X}$ has i,j element $\left(b^{i+j}-a^{i+j}\right)/(i+j)$. Since $\int_a^b [g(x|\boldsymbol{\alpha})]^2\,dx > 0$ for all $\boldsymbol{\alpha}$ then the matrix X must be positive definite. Next we notice that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}[g(x_i|\boldsymbol{\alpha})]^2 &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\alpha}^T\boldsymbol{X}_i\boldsymbol{\alpha}\\ &= \boldsymbol{\alpha}^T\left(\frac{1}{n}\sum\boldsymbol{X}_i\right)\boldsymbol{\alpha}\\ &= \boldsymbol{\alpha}^T\boldsymbol{X}_n\boldsymbol{\alpha}\end{aligned}$$

and therefore $\boldsymbol{X}_n \to \boldsymbol{X}$ and therefore, denoting the eigen-values of $\boldsymbol{X}_n$ as $\boldsymbol{\lambda}_n$ and the eigenvalues of $\boldsymbol{X}$ as $\boldsymbol{\lambda}$, we have $\boldsymbol{\lambda}_n \to \boldsymbol{\lambda}$

Letting $\lambda_{n,m}$ and $\lambda_{n,M}$ be the minimum and maximum eigen-values of $\boldsymbol{X}_n$ be the largest, then $\lambda_{n,m}^2\|\boldsymbol{\alpha}\|^2 \le \frac{1}{n}\sum[g(x|\boldsymbol{\alpha})]^2 \le \lambda_{n,M}^2\|\boldsymbol{\alpha}\|^2$. □

The values $\lambda_{n,m}, \lambda_{n,M}$ depend on the interval that the polynomial is integrated/summed over and that if $a=b$, then the integral is zero. In the following lemmas, we assume that there is some minimal distance between two knot-points and between a knot-point and the boundary values a, b.

Lemma 2. *Given a degree p spline $g(x|\boldsymbol{\theta})$ with κ knot points on $[a,b]$, let $\tau = (|a|\vee|b|)^\kappa$. Then $\forall\,\delta > 2\tau$, $\exists\,\lambda_n > 0$ such that if $\|\boldsymbol{\theta}\| > \delta$ then $\frac{1}{n}\sum[g(x_i|\boldsymbol{\theta})]^2 > \left(\delta^2+\tau^2\right)\lambda_n$.*

Proof. Notice that $||\boldsymbol{\theta}||^2 > \delta^2 > 4\tau^2$ implies $||\boldsymbol{\alpha}||^2 > \delta^2-\tau^2$. First we consider the case of $\kappa=1$. If $\alpha_0^2+\cdots+\alpha_p^2 > \left(\delta^2+\tau^2\right)/9$ then $\frac{1}{n}\sum[g(x_i|\boldsymbol{\theta})]^2\,1_{[a,t]}(x_i) > \lambda_n\left(\delta^2+\tau^2\right)$ for some $\lambda_n > 0$. If $\alpha_0^2+\cdots+\alpha_p^2 \le \left(\delta^2+\tau^2\right)/9$ then $\alpha_{p+1}^2 \ge 3\left(\delta^2+\tau^2\right)/4$. Therefore $(\alpha_p+\alpha_{p+1})$, the coefficient of the x^p term of the polynomial on $[t_1,b]$ is

$$\begin{aligned}\|\alpha_p+\alpha_{p+1}\|^2 &> \|\alpha_{p+1}\|^2-\|\alpha_p\|^2\\ &> \frac{3\left(\delta^2+\tau^2\right)}{4}-\frac{\left(\delta^2+\tau^2\right)}{4}\\ &> \frac{1}{2}\left(\delta^2+\tau^2\right)\end{aligned}$$

and thus the squared norm of the coefficients of the polynomial on $[t_1,b]$ must also be greater than $\frac{1}{2}\left(\delta^2+\tau^2\right)$ and thus $\frac{1}{n}\sum[g(x_i|\boldsymbol{\theta})]^2\,1_{[t,b]}(x_i) > \lambda_n\left(\delta^2+\tau^2\right)$ for some $\lambda_n > 0$. The proof for multiple knots is similar, only examining all $\kappa+1$ polynomial sections for one with coefficients with squared norm larger than some fraction of $\left(\delta^2+\tau^2\right)$. □

Lemma 3. *For all $\delta > 0$, there exists $\lambda_n > 0$ such that for all $\boldsymbol{\theta}\notin B(\boldsymbol{\theta}_0,\delta)$ then $\frac{1}{n}\sum\left(g(\boldsymbol{\theta}_0,\boldsymbol{\theta},x_i)\right)^2 > \lambda_n\delta$.*

Proof. By the previous lemma, for all $\Delta > 2\tau$ there exists $\exists\,\Lambda_n > 0$ such that for all $\boldsymbol{\theta}\notin B(\boldsymbol{\theta}_0,\Delta)$ then $\frac{1}{n}\sum\left(g(\boldsymbol{\theta}_0,\boldsymbol{\theta},x_i)\right)^2 > \Lambda_n\Delta$. We now consider the region

$$\mathcal{C} = \text{closure}\left[B\left(\boldsymbol{\theta}_0,\Delta\right)nB\left(\boldsymbol{\theta}_0,\delta\right)\right]$$

Assume to the contrary that there exists $\delta > 0$ such that $\forall \lambda_n > 0$, $\exists \boldsymbol{\theta} \in \mathcal{C}$ such that $\frac{1}{n}\sum (g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i))^2 \leq \lambda_n \delta$ and we will seek a contradiction. By the negation, there exists a sequence $\boldsymbol{\theta}_n \in \mathcal{C}$ such that $\frac{1}{n}\sum (g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i))^2 \to 0$. But since $\boldsymbol{\theta}_n$ is in a compact space, there exists a sub-sequence $\boldsymbol{\theta}_{n_k}$ that converges to $\boldsymbol{\theta}_\infty \in \mathcal{C}$ and $\frac{1}{n}\sum (g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i))^2 = 0$. But since $\boldsymbol{\theta}_0 \notin \mathcal{C}$ this is a contradiction. □

Corollary 4. *There exists* λ *such that for any* $\delta > 0$ *and* $\boldsymbol{\theta} \notin B(\boldsymbol{\theta}_0, \delta)$

$$\frac{1}{n}\sum_{i=1}^{n} \left[g\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i\right) + \sigma_0 \epsilon_i\right]^2 \quad \geq \quad \lambda_n^2 \delta^2 + O_p\left(n^{-1/2}\right) + \sigma_0^2.$$

We now focus our attention on the ratio of the maximum value of a polynomial and its integral.

Lemma 5. *Given a degree* p *polynomial* $g\left(x|\boldsymbol{\alpha}\right)$ *on* $[a, b]$*, then*

$$\frac{\max_{i\in\{1,\dots,n\}} \left[g\left(x_i|\boldsymbol{\alpha}\right)\right]^2}{\frac{1}{n}\sum_{i=1}^{n} \left[g\left(x_i|\boldsymbol{\alpha}\right)\right]^2 \, dx} \leq \frac{\lambda_M^2}{\lambda_{n,m}^2} \to \frac{\lambda_M^2}{\lambda_m^2}$$

for some $\lambda_M, \lambda_m > 0$.

Proof. Since we can write $\left[g\left(x|\boldsymbol{\alpha}\right)\right]^2 = \boldsymbol{\alpha}^T W_x \boldsymbol{\alpha}$ for some non-negative definite matrix W_x which has a maximum eigen-value $\lambda_{M,x}$, and because the the maximum eigen-value is a continuous function in x, let $\lambda_M = \sup \lambda_{M,x}$. Then the maximum of $\left[g\left(x|\boldsymbol{\alpha}\right)\right]^2$ over $x \in [a, b]$ is less than $\left\|\boldsymbol{\alpha}\right\|^2 \lambda_M^2$. The denominator is bounded from below by $\left\|\boldsymbol{\alpha}\right\|^2 \lambda_{n,m}^2$. □

Lemma 6. *Given a degree* p *spline* $g\left(x|\boldsymbol{\theta}\right)$ *on* $[a, b]$*, then*

$$\frac{\max \left[g\left(x|\boldsymbol{\theta}\right)\right]^2}{\int_a^b \left[g\left(x|\boldsymbol{\theta}\right)\right]^2 \, dx} \leq \frac{\lambda_M^2}{\lambda_m^2}$$

for some $\lambda_M, \lambda_m > 0$.

Proof. Since a degree p spline is a degree p polynomial on different regions defined by the knot-points, and because the integral over the whole interval $[a, b]$ is greater than the integral over the regions defined by the knot-points, we can use the previous lemma on each section and then chose the largest ratio. □

Lemma 7. *Given a degree* p *spline* $g\left(x|\boldsymbol{\theta}\right)$ *on* $[a, b]$ *then*

$$\frac{n^{-1/2} \max_i \left[\epsilon_i \sigma_0 + g\left(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i\right)\right]^2}{n^{-1}\sum_{i=1}^{n} \left[\epsilon_i \sigma_0 + g\left(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i\right)\right]^2} = O_p\left(1\right) \tag{1}$$

uniformly over $\boldsymbol{\theta}$.

Proof. Notice

$$\begin{aligned}\frac{n^{-1/2}\max_i\left[\epsilon_i\sigma_0+g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)\right]^2}{n^{-1}\sum_{i=1}^n\left[\epsilon_i\sigma_0+g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)\right]^2} &\leq \frac{2n^{-1/2}\max_i\left[\epsilon_i^2\sigma_0^2\right]+2n^{-1/2}\max_i\left[g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)\right]^2}{n^{-1}\sum_{i=1}^n\left[\epsilon_i\sigma_0+g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)\right]^2}\\ &= \frac{2\sigma_0^2 n^{-1/2}\max_i\epsilon_i^2+\max_i\left[g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)\right]^2}{n^{-1}\sum_{i=1}^n\left[\epsilon_i\sigma_0+g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)\right]^2}\\ &= \frac{O_p\left(\frac{\log n}{\sqrt{n}}\right)+\max_i\left[g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)\right]^2}{n^{-1}\sum_{i=1}^n\left[\epsilon_i\sigma_0+g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)\right]^2}\end{aligned}$$

and since $n^{-1}\sum_{i=1}^n\left[\epsilon_i\sigma_0+g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)\right]^2 \xrightarrow{P} \frac{1}{b-a}\int_a^b\left(g(\boldsymbol{\theta}_0,\boldsymbol{\theta},x)\right)^2 dx+\sigma_0^2$, and lemma 8 bounds the ratio of the terms that involve $\boldsymbol{\theta}$, this ratio is bounded in probability uniformly over $\boldsymbol{\theta}$. □

B.5. Assumptions B1

Returning to assumption B1, we now consider $\boldsymbol{\xi}\notin B(\boldsymbol{\xi}_0,\delta)$ and

$$\begin{aligned}L_n\left(\boldsymbol{\xi}\right) &= \sum\log\left\{\frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{-1}{2\sigma}\sum\left(y_i-N(x_i,\boldsymbol{t})^T\boldsymbol{\alpha}\right)^2\right]\right\}\\ &= -\frac{n}{2}\log\left(2\pi\right)-n\log\sigma-\frac{1}{2\sigma}\sum\left[y_i-N(x_i,\boldsymbol{t})^T\boldsymbol{\alpha}\right]^2\\ &= -\frac{n}{2}\log\left(2\pi\right)-n\log\sigma-\frac{1}{2\sigma}\sum\left[N(x_i,\boldsymbol{t}_0)^T\boldsymbol{\alpha}_0+\sigma_0\epsilon_i-N(x_i,\boldsymbol{t})^T\boldsymbol{\alpha}\right]^2\\ &= -\frac{n}{2}\log\left(2\pi\right)-n\log\sigma-\frac{1}{2\sigma}\sum\left[g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)+\sigma_0\epsilon_i\right]^2\end{aligned}$$

and therefore

$$\begin{aligned}&\frac{1}{n}\left(L_n(\boldsymbol{\xi})-L_n(\boldsymbol{\xi}_0)\right)\\ &\quad= -\log\sigma-\frac{1}{2n\sigma^2}\sum\left[g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)+\sigma_0\epsilon_i\right]^2+\log\sigma_0+\frac{1}{2n\sigma_0}\sum\left[g\left(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0,x_i\right)+\sigma_0\epsilon_i\right]^2\\ &\quad= \log\frac{\sigma_0}{\sigma}-\frac{1}{2n\sigma^2}\sum\left[g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)+\sigma_0\epsilon_i\right]^2+\frac{1}{2n\sigma_0^2}\sum\left[\sigma_0\epsilon_i\right]^2\\ &\quad= \log\frac{\sigma_0}{\sigma}-\frac{\left(\lambda_n\left(\boldsymbol{\theta},\boldsymbol{\theta}_0\right)\right)^2}{2\sigma^2}-\frac{\sigma_0^2}{2\sigma^2}+\frac{1}{2n}\sum\left[\epsilon_i\right]^2\end{aligned}$$

where

$$\left[\lambda_n\left(\boldsymbol{\theta},\boldsymbol{\theta}_0\right)\right]^2 = \frac{1}{n}\sum\left[g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x_i\right)+\sigma_0\epsilon_i\right]^2-\sigma_0^2$$

which converges in probability to $\frac{1}{b-a}\int_a^b\left[g\left(\boldsymbol{\theta},\boldsymbol{\theta}_0,x\right)\right]^2dx$. The function goes to $-\infty$ as $\sigma\to 0$ and $\sigma\to\infty$. Taking the derivative

$$\frac{d}{d\sigma}\left[\log\frac{\sigma_0}{\sigma}-\frac{1}{2\sigma^2}\left[\left(\lambda_n\right)^2+\sigma_0^2\right]+\frac{1}{2n}\sum\epsilon_i^2\right]=-\frac{1}{\sigma}+\frac{1}{\sigma^3}\left[\left(\lambda_n\right)^2+\sigma_0^2\right]$$

and setting it equal to zero yields a single critical point of at $\sigma^2 = \left[(\lambda_n)^2 + \sigma_0^2\right]$ which results in a maximum of

$$\log\left(\frac{\sigma_0}{\sqrt{(\lambda_n)^2 + \sigma_0^2}}\right) - \frac{1}{2} + \frac{1}{2}n^{-1}\sum \epsilon_i^2 \tag{2}$$

which bounded away from zero in probability for $\boldsymbol{\xi} \notin B(\boldsymbol{\xi}_0, \delta)$

B.6. Assumption C1

Assumption C1 is

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

First notice

$$\begin{aligned}
L(\boldsymbol{\xi}, Y_i) &= -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2}\left(Y_i - N(x_i, \boldsymbol{t})^T\boldsymbol{\alpha}\right)^2 \\
&= -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2}\left(\epsilon_i\sigma_0 + N(x_i, \boldsymbol{t}_0)^T\boldsymbol{\alpha}_0 - N(x_i, \boldsymbol{t})^T\boldsymbol{\alpha}\right)^2 \\
&= -\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2}\left(\epsilon_i\sigma_0 + g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i)\right)^2
\end{aligned}$$

and we consider $\mathcal{C} = \{\boldsymbol{\xi} : \boldsymbol{\xi} \notin B(\boldsymbol{\xi}_0, \delta)\}$. Define

$$\begin{aligned}
f_n(\boldsymbol{\xi}) &= \frac{\min\ L(\boldsymbol{\xi}, Y_i)}{|L_n(\boldsymbol{\xi}) - L_n(\boldsymbol{\xi}_0)|} \\
&= \frac{-\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2}\max\left[\epsilon_i\sigma_0 + g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i)\right]^2}{n\cdot\frac{1}{n}|L_n(\boldsymbol{\xi}) - L_n(\boldsymbol{\xi}_0)|}
\end{aligned}$$

and notice that the denominator is bounded away from 0 by 2.

$$\begin{aligned}
f_n(\boldsymbol{\xi}) &= \frac{-\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2}\max\left[\epsilon_i\sigma_0 + g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i)\right]^2}{-n\cdot\frac{1}{n}\left(L_n(\boldsymbol{\xi}) - L_n(\boldsymbol{\xi}_0)\right)} \\
&= \frac{\frac{1}{\sqrt{n}}\left[-\frac{1}{2}\log(2\pi) - \log\sigma - \frac{1}{2\sigma^2}\max\left[\epsilon_i\sigma_0 + g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i)\right]^2\right]}{-\sqrt{n}\cdot\frac{1}{n}\left[n\log\frac{\sigma_0}{\sigma} - \frac{1}{2\sigma^2}\sum\left[g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i) + \sigma_0\epsilon_i\right]^2 + \frac{1}{2}\sum\epsilon_i^2\right]} \\
&= \frac{1}{\sqrt{n}}\cdot\frac{-\frac{1}{2\sqrt{n}}\log(2\pi) - \frac{1}{\sqrt{n}}\log\sigma - \frac{1}{2\sqrt{n}\sigma^2}\max\left[\epsilon_i\sigma_0 + g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i)\right]^2}{-\log\frac{\sigma_0}{\sigma} + \frac{1}{2n\sigma^2}\sum\left[g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i) + \sigma_0\epsilon_i\right]^2 - \frac{1}{2n}\sum\epsilon_i^2} \\
&= \frac{1}{\sqrt{n}}\left[\frac{-\frac{1}{2\sqrt{n}}\log(2\pi)}{-\log\frac{\sigma_0}{\sigma} + \frac{1}{2n\sigma^2}\sum\left[g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i) + \sigma_0\epsilon_i\right]^2 - \frac{1}{2n}\sum\epsilon_i^2} + \right. \\
&\qquad \left.\frac{-\frac{1}{\sqrt{n}}\log\sigma - \frac{1}{2\sqrt{n}\sigma^2}\max\left[\epsilon_i\sigma_0 + g(\boldsymbol{\theta}_0, \boldsymbol{\theta}, x_i)\right]^2}{-\log\frac{\sigma_0}{\sigma} + \frac{1}{2n\sigma^2}\sum\left[g(\boldsymbol{\theta}, \boldsymbol{\theta}_0, x_i) + \sigma_0\epsilon_i\right]^2 - \frac{1}{2n}\sum\epsilon_i^2}\right]
\end{aligned}$$

We consider the infimums of the terms inside the brackets separately.

For the first term, since the denominator is bounded in probability above 0 uniformly in $\boldsymbol{\theta}$, and the numerator goes to zero, the infimum of the first term goes to 0 in probability.

The second term is uniformly bounded over $\boldsymbol{\theta}$ by lemma 9. Notice that the numerator is

$$\begin{aligned}
-\frac{1}{\sqrt{n}}\log\sigma - \frac{1}{2\sqrt{n}\sigma^2}\max\left[\epsilon_i\sigma_0 + g(\boldsymbol{\theta}_0,\boldsymbol{\theta},x_i)\right]^2 & \\
\geq \quad -\frac{1}{\sqrt{n}}\log\sigma - \frac{\max\left[\epsilon_i\sigma_0\right]^2}{\sqrt{n}\sigma^2} - \frac{\max\left[g(\boldsymbol{\theta}_0,\boldsymbol{\theta},x_i)\right]^2}{\sqrt{n}\sigma^2} & \\
= \quad -\frac{1}{\sqrt{n}}\log\sigma - \frac{\sigma_0^2\,O_p\left(\log\, n\right)}{\sqrt{n}\sigma^2} - \frac{\max\left[g(\boldsymbol{\theta}_0,\boldsymbol{\theta},x_i)\right]^2}{\sqrt{n}\sigma^2} & \\
\geq \quad \frac{-\log n}{\sqrt{n}}\,\log\sigma - \frac{\sigma_0^2\,O_p\left(\log\, n\right)}{\sqrt{n}\sigma^2} - \frac{\max\left[g(\boldsymbol{\theta}_0,\boldsymbol{\theta},x_i)\right]^2}{\sqrt{n}\sigma^2} &
\end{aligned}$$

and all three terms of the numerator converge to 0 for every σ. Therefore for $\sigma\in[0,d]$ for some large d, the infimum converges to 0. For $\sigma>d$, the $\log\sigma$ terms dominate and the infimum occurs at $\sigma=d$ which also converges to 0. Therefore

$$\inf_{\boldsymbol{\xi}\notin B(\boldsymbol{\xi}_0,\delta)}\frac{\min L\left(\boldsymbol{\xi},Y_i\right)}{\left|L_n\left(\boldsymbol{\xi}\right)-L_n\left(\boldsymbol{\xi}_0\right)\right|}\xrightarrow{P}0.$$

B.7. Assumptions C2

Finally we turn our attention to the Jacobian. Recall that the Jacobian is

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

where

$$\boldsymbol{B_\alpha}=\left[\begin{array}{ccccccc}1 & x_{(1)} & \dots & x_{(1)}^p & (x_{(1)}-t_1)_+^p & \dots & (x_{(1)}-t_\kappa)_+^p\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 1 & x_{(l)} & \dots & x_{(l)}^p & (x_{(l)}-t_1)_+^p & \dots & (x_{(l)}-t_\kappa)_+^p\end{array}\right],$$

$$\boldsymbol{B_t}=\left[\begin{array}{ccc}\alpha_{1+p+1}\left(x_{(1)}-t_1\right)_+^{p-1}I\left(x_{(1)}-t_1\right) & \dots & \alpha_{1+p+\kappa}\left(x_{(1)}-t_\kappa\right)_+^{p-1}I\left(x_{(1)}-t_\kappa\right)\\ \vdots & \ddots & \vdots\\ \alpha_{1+p+1}\left(x_{(l)}-t_1\right)_+^{p-1}I\left(x_{(l)}-t_1\right) & \dots & \alpha_{1+p+\kappa}\left(x_{(l)}-t_\kappa\right)_+^{p-1}I\left(x_{(l)}-t_\kappa\right)\end{array}\right],$$

and

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

Following the notation of Yeo and Johnson, we suppress parenthesis and 0 subscripts. We consider the $\boldsymbol{\xi}$ in compact space $\bar{B}(\boldsymbol{\xi}_0, \delta)$. We notice that for $\delta < \sigma^{-2}$ that $J(\boldsymbol{y}; \boldsymbol{\xi}) \leq \delta^{\kappa+1} p^{\kappa} g(\boldsymbol{y})$ for some $g(\boldsymbol{y})$ because $\boldsymbol{B_\alpha}$ and $\boldsymbol{B_t}$ are functions of $\boldsymbol{x}, \boldsymbol{t}$ which are bounded.

We let S_M^l be the unit square in $\mathbb{R}^l$ of radius M.

Finally we notice that $J_j(y_1, \ldots, y_j; \boldsymbol{\xi}) = E\left[J\left(y_1, \ldots, y_j, Y_{j+1}, \ldots, Y_l; \boldsymbol{\xi}\right)\right]$ is a polynomial in $\boldsymbol{\theta}$ scaled by σ^2, which is equicontinuous on compacts of $\boldsymbol{\xi}$ where σ is bounded away from 0.

Appendix C: Full simulation results

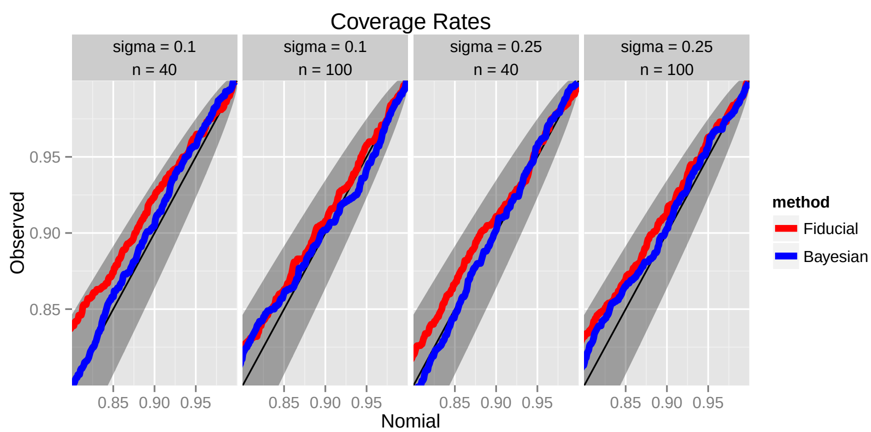

FIG 1. *Coverage rates for the single knot scenario. The color (red, blue) represents the method (fiducial, Bayesian).*

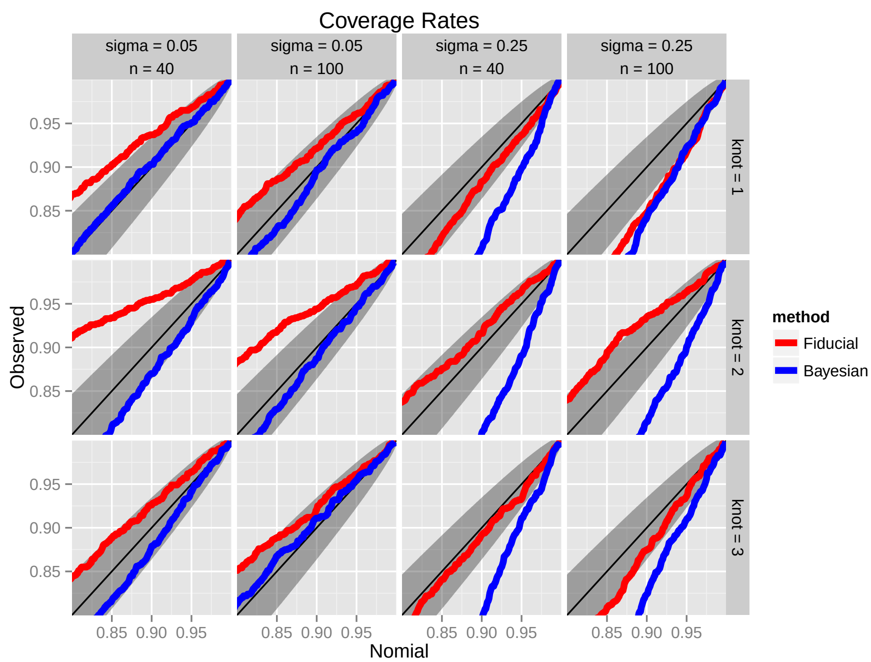

FIG 2. *Coverage rates for the three knot "Simple" scenario. The color (red, blue) represents the method (fiducial, Bayesian). The topmost panel is the coverage of knot 1 in the sigma =* $0.1, n = 40$ *simulation.*

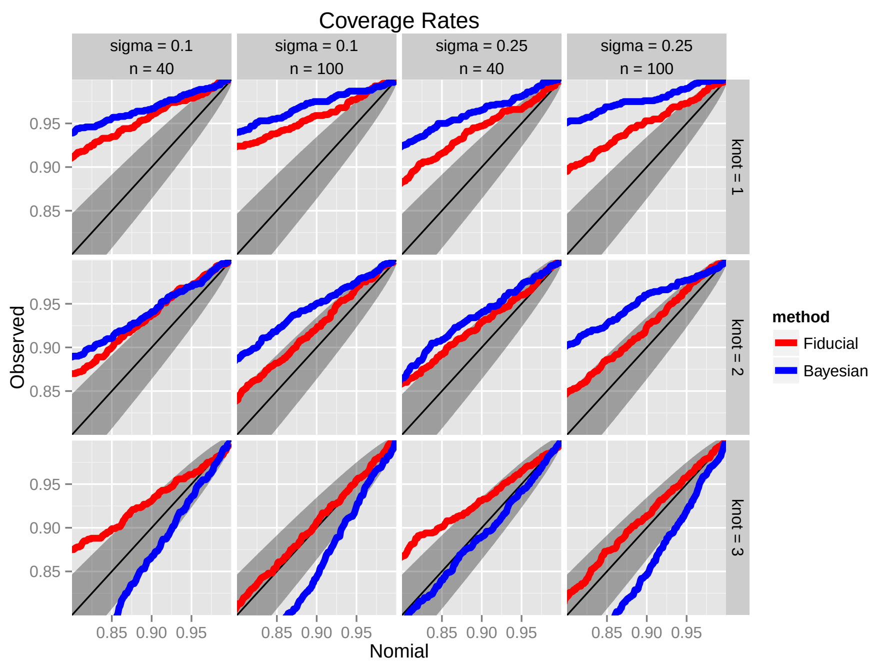

FIG 3. *Coverage rates for the three knot "Clustered" scenario. The color (red, blue) represents the method (fiducial, Bayesian). The topmost panel is the coverage of knot 1 in the sigma* $= 0.1, n = 40$ *simulation.*

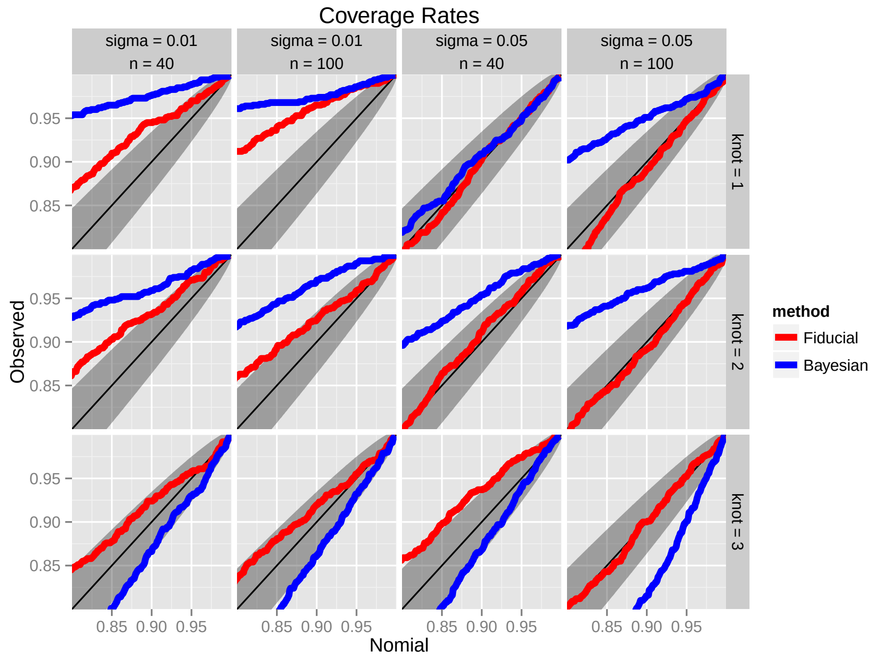

FIG 4. *Coverage rates for the three knot "Subtle" scenario. The color (red, blue) represents the method (fiducial, Bayesian). The topmost panel is the coverage of knot 1 in the sigma =* $0.1, n = 40$ *simulation.*

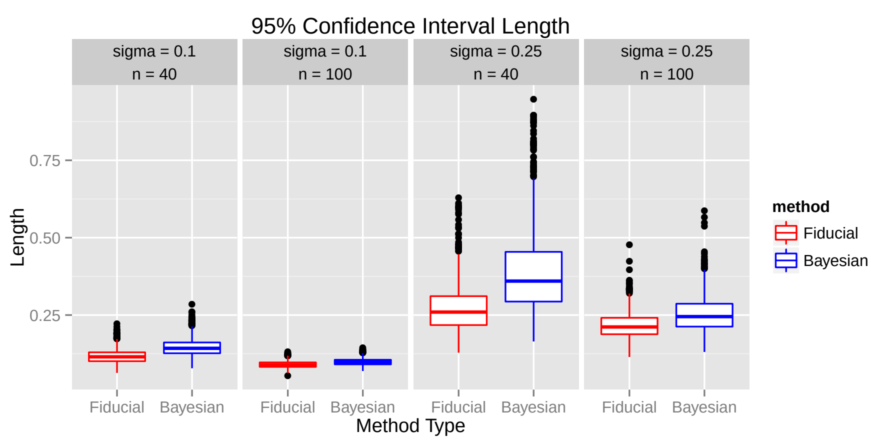

FIG 5. *Confidence interval lengths for the single knot scenario. The color (red, blue) represents the method (fiducial, Bayesian).*

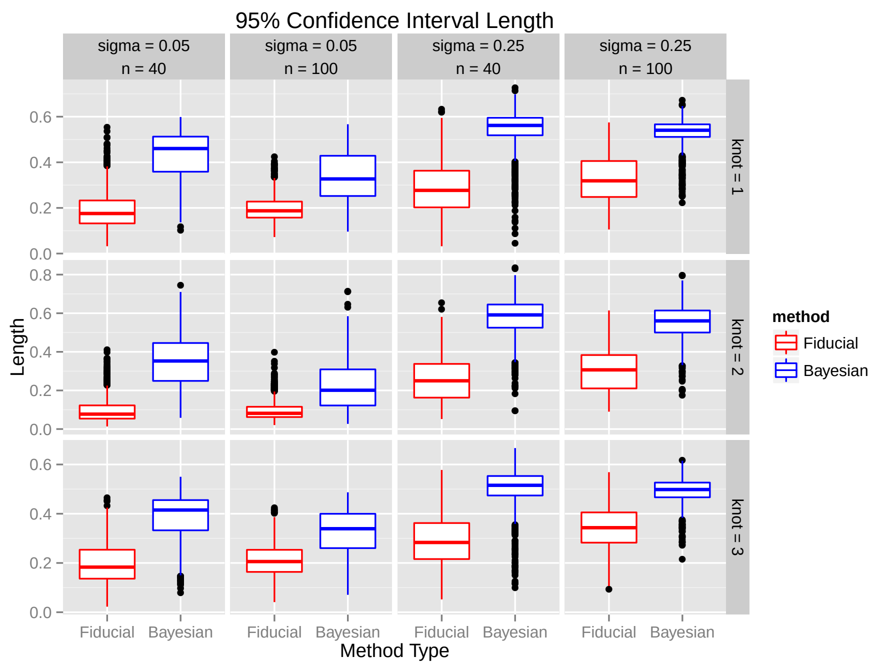

FIG 6. *Confidence interval lengths for the three knot "Simple" scenario. The color (red, blue) represents the method (fiducial, Bayesian). The topmost panel is the coverage of knot 1 in the sigma* $= 0.1, n = 40$ *simulation.*

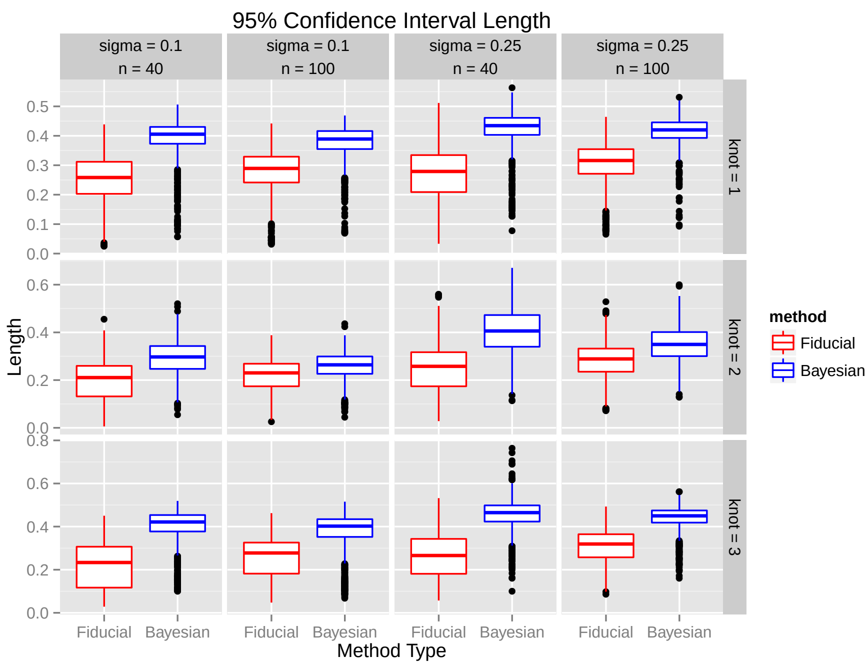

FIG 7. *Confidence interval lengths for the three knot "Clustered" scenario. The color (red, blue) represents the method (fiducial, Bayesian). The topmost panel is the coverage of knot 1 in the sigma* $= 0.1, n = 40$ *simulation.*

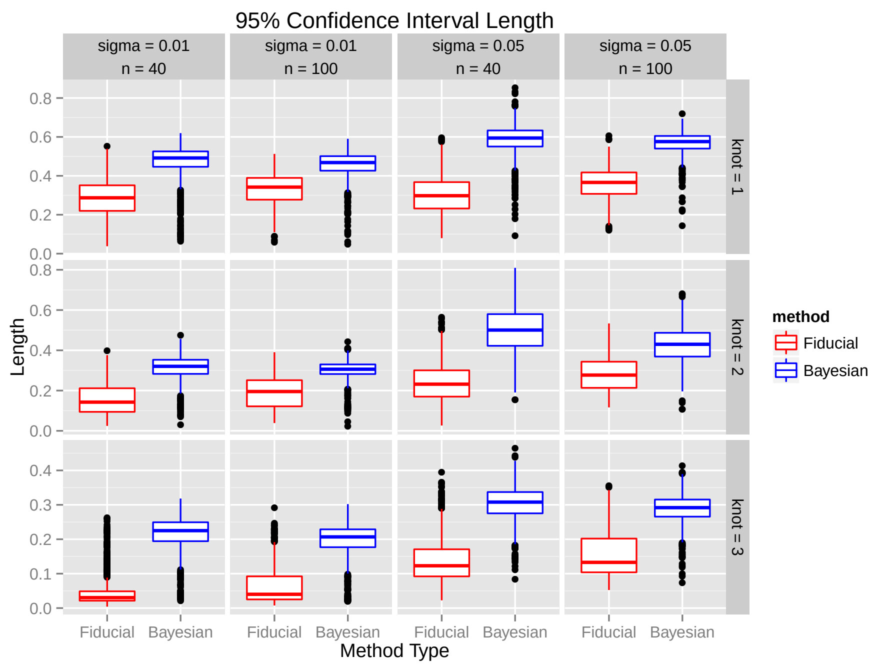

FIG 8. *Confidence interval lengths for the three knot "Subtle" scenario. The color (red, blue) represents the method (fiducial, Bayesian). The topmost panel is the coverage of knot 1 in the sigma* $= 0.1, n = 40$ *simulation.*