Simultaneous confidence bands and the volume-of-tube method

Xiaolei LU

Doctor of Philosophy

Department of Statistical Science
School of Multidisciplinary Sciences
SOKENDAI (The Graduate University for Advanced Studies)

Simultaneous confidence bands and the volume-of-tube method

A DISSERTATION
SUBMITTED TO THE FACULTY OF
THE SCHOOL OF MULTIDISCIPLINARY SCIENCES
THE DEPARTMENT OF STATISTICAL SCIENCE
THE GRADUATE UNIVERSITY FOR ADVANCED STUDIES
BY

Xiaolei LU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Satoshi Kuriki, Advisor

March 2017

Acknowledgements

I am grateful to my supervisor, Satoshi Kuriki, who has taken much time to teach me many things during my Ph.D. course. I respect his knowledge, kindness, and wisdom very much. He is one of the best teachers in my life. I also want to thank my former supervisor, Wei Gao, whose spirit will affect me all my life. I also thank Fukumizu Kenji and Hironori Fujisawa, who provided many useful comments and suggestions. A big thanks to Michael S. Waterman, Haiyan Huang, Lucy Yin, Amy, Long Hu, and Hu Ding, who greatly supported me in the U.S. Thanks to their kindness, I will never forget them. I very much appreciate Yuhong Wang and Xia Ying, whose great ideas and beautiful hearts warm me and brighten me up. I feel very lucky to have met Higuchi Sensei, Tamura Sensei, Hasegawa Sensei, Kitagawa Sensei, Hasebe Sensei, Zhuang Sensei, and Notsu Sensei. I am thankful for their kindness. Thanks to my friends Hua Wang, Haiyan Nie, and Shuyun Xu, who are always with me. Thanks to all my family. I am always like a child in your eyes, but I would like to grow up from now on. Thank you for giving me a happy life.

Abstract

This study focuses on simultaneous confidence bands and the volume-of-tube method. Simultaneous confidence bands have been used in various statistical problems. The volume-of-tube method can be used in the construction of simultaneous confidence bands.

The problem concerning the construction of simultaneous confidence bands in a regression model originates with Working and Hotelling (1929). They formalized this problem as the construction of confidence intervals for an estimated regression line, and provided a critical value by making use of the Cauchy-Schwarz inequality. Subsequently, many reports concerning the relaxation of these conditions have appeared.

In the case of one regression model, Wynn and Bloomfield (1971) pointed out that the use of the Cauchy-Schwarz inequality leads to conservative bands for simple regression with unrestricted domain of the explanatory variables. Uusipaikka (1983) constructed exact confidence bands for linear regression when $\mathcal{X}$ is a finite interval.

In the case of the linear regression models, there are a lot of research in literature. For example, simultaneous confidence intervals are used in Scheffé (1953) to assess any contrasts between several normal means. In this study, the problem of assessing any contrasts between several simple linear regression models is considered by using simultaneous confidence bands. Using numerical integration, Spurrier (1999) constructed exact simultaneous confidence bands for all of the contrasts between several regression lines over the whole range $(-\infty, \infty)$ of the explanatory variable when the design matrices of the regression lines are all equal. Jamshidian, Liu, and Bretz (2010) proposed a simulation-based method to construct simultaneous confidence bands for all of the contrasts between the linear regression models when the explanatory variable is restricted to an interval and the design matrices of the regression lines may be different. Naiman (1986) gives a method for constructing conservative Scheff $\acute{e}$-type simultaneous confidence bands for a single curvilinear regression model over finite intervals. Unlike these studies, we consider constructing simultaneous confidence bands for all of the contrasts between several nonlinear regression models. The tube formula is given in a mathematical form via the volume-of-tube method.

The chapters are arranged as follows. We provide a brief review of multiple regression models in Chapter 1. Chapter 2 summarizes simultaneous confidence bands for simple and multiple regression models, and we then address the problem of the construction of simultaneous confidence bands for all of the contrasts between several nonlinear regression models. We propose simultaneous confidence bands of the hyperbolic type for the contrasts between several nonlinear (curvilinear) regression

curves. Chapter 3 looks at the volume-of-tube method. We give the definition of the tube and critical radius, and then we summarize the volume-of-tube method for evaluating the upper tail probability. In addition, we discuss the expectation of the Euler-Poincaré characteristic heuristic. Moreover, we prove that the formula obtained is equivalent to the expectation of the Euler-Poincaré characteristic of the excursion set of the chi-square random process and, hence, is conservative. Using this result, Takemura and Kuriki (2002) provide an alternative proof that the confidence band of Naiman (1986) is conservative. Chapter 4 uses the volume-of-tube method to derive an upper tail probability formula for the maximum of a chi-square random process, which is sufficiently accurate in commonly used tail regions. The critical value of a confidence band is determined from the distribution of the maximum of a chi-square random process defined on the domain of the explanatory variables. The tube formula is given in a mathematical form. We prove that the simultaneous confidence bands we propose are conservative. This result is therefore a generalization of Naiman's inequality for Gaussian random processes. In order to test our method, we give a numerical example to determine the accuracy of the approximation formula we propose, which further demonstrate that the confidence bands obtained by the tube method are always conservative and very accurate. To investigate what happens under model misspecification, we conduct a Monte Carlo simulation study. The study shows, too small of a model should surely be avoided, whereas, a larger model has the disadvantage of having a wider confidence band.

As an illustrative example, the growth curves of consomic mice are analyzed in Chapter 5. A study under model misspecification is also conducted in the application. Chapter 6 considers the statistical parametric mapping approach as future work. Details of the proofs are in the Appendix.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Multiple regression analysis is a powerful technique used for predicting the relationship between a continuous random variable Y and several independent variables. Let $Y_1, Y_2, \ldots, Y_p$ be a set of predictors believed to be related to a response variable Y. Let $Y = (Y_1, \ldots, Y_n)^\top$ and $x_i = (x_{1i}, \ldots, x_{ni})^\top$ be the sequences of observations that follow the regression model.

$$Y_j = \beta_0 + \beta_1 x_{j1} + \ldots + \beta_p x_{jp} + \varepsilon_j, \quad j = 1, \ldots, n,$$

where $\beta_i, i = 0, 1, \ldots, p$ are unknown regression coefficients and ε_j represents mutually independent $\mathcal{N}(0, \sigma^2)$ random variables. We rewrite the regression model in matrix form as

$$Y = X\beta + \varepsilon,$$

where $X = (\mathbb{1}, x_1, \ldots, x_p)$, $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^\top$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, and $\mathbb{1}$ is a column vector of size n with all elements equal to one. The matrix X is defined as a design matrix. Without loss of generality, we assume that X is of full column rank.

For more details, see Anderson (2009).

1.1 Parameter estimation

The method of least squares estimation is a standard approach to estimating β in regression analysis. We can obtain the least squares estimator $\widehat{\beta}$ of β by minimizing the least squares criterion, as in Liu (2010), given by

$$L(\beta) = ||Y - X\beta||^2 = (Y - X\beta)^\top (Y - X\beta).$$

Thus, the least squares estimator must satisfy

$$\frac{\partial L(\beta)}{\partial \beta}\Big|_{\beta=\widehat{\beta}} = -2X^\top Y + 2X^\top X\widehat{\beta} = 0.$$

Because X is of full column rank, $X^\top X$ is non-singular, and the normal equation leads to the least squares estimator

$$\widehat{\beta} = (X^\top X)^{-1}X^\top Y.$$

Fitting the model, we can obtain

$$\widehat{Y} = X\widehat{\beta} = X(X^\top X)^{-1}X^\top Y = HY,$$

where $H = X(X^\top X)^{-1}X^\top$ is called the hat matrix such that $H(I-H) = 0$ since $H^2 = H$.

The vector of residuals is defined as

$$\widehat{\varepsilon} = Y - \widehat{Y} = (I-H)Y.$$

The estimator $\widehat{\sigma}^2$ of σ^2 is defined as

$$\widehat{\sigma}^2 = ||\widehat{\varepsilon}||^2/(n-p-1).$$

Because y is a realization of a random vector Y with $\mathrm{E}(Y) = X\beta$, we obtain the following theorem.

Theorem 1.1.1. *Under the standard normality assumptions, we have the following properties.*

(i) $\widehat{\beta} \sim \mathcal{N}_{p+1}(\beta, \sigma^2(X^\top X)^{-1})$.
(ii) $\widehat{\varepsilon} \sim \mathcal{N}_n(0, \sigma^2(I-H))$.
(iii) $\widehat{\sigma}^2 \sim \frac{\sigma^2}{n-p-1}\chi^2_{n-p-1}$.
(iv) $\widehat{\beta}$ *and* $\widehat{\varepsilon}$ *are independent.*
(v) $\widehat{\beta}$ *and* $\widehat{\sigma}^2$ *are independent.*

1.2 Confidence intervals

It is clear that $x^\top\beta$ can be estimated by $x^\top\widehat{\beta}$, since

$$x^\top(\widehat{\beta}-\beta) \sim \mathcal{N}(0, \sigma^2 x^\top(X^\top X)^{-1}x).$$

When σ is known,

$$\frac{x^\top(\widehat{\beta}-\beta)}{\sigma\sqrt{x^\top(X^\top X)^{-1}x}}$$

follows a normal distribution. Hence, a $1-\alpha$ confidence interval for $x^\top\beta$ is given by

$$\Pr\left\{x^\top\beta \in x^\top\widehat{\beta} \pm Z_{\alpha/2}\sigma\sqrt{x^\top(X^\top X)^{-1}x}\right\} = 1-\alpha,$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ point of the normal distribution.

When σ is unknown, it follows from Theorem 1.1.1 that β is independent of $\widehat{\sigma}$.

$$\frac{x^\top(\widehat{\beta}-\beta)}{\widehat{\sigma}\sqrt{x^\top(X^\top X)^{-1}x}}$$

follows a t distribution with $n-p-1$ degrees of freedom. Hence, a $1-\alpha$ confidence interval for $x^\top\beta$ is given by

$$\Pr\left\{x^\top\beta \in x^\top\widehat{\beta} \pm t_{\alpha/2}\widehat{\sigma}\sqrt{x^\top(X^\top X)^{-1}x}\right\} = 1-\alpha,$$

where $t_{\alpha/2}$ is the upper $\alpha/2$ point of the t distribution with $n-p-1$ degrees of freedom, as in Liu (2010).

1.3 The layout of this thesis

The following is a brief outline of this thesis. Chapter 1 provides a brief review of multiple regression models. In Chapter 2, we review simultaneous confidence bands for simple and multiple regression models, and we then address the problem of the construction of simultaneous confidence bands for nonlinear regression models. Chapter 3 looks at the volume-of-tube method, and we summarize the volume-of-tube method for evaluating the upper tail probability of the maximum of a Gaussian random field. The volume-of-tube method and its related method, referred to as the expected Euler-characteristic heuristic, are briefly summarized in Chapter 3. In Chapter 4, we define a Gaussian random field and a chi-square random process as pivotal quantities. We show that the critical value is determined from the upper tail probability of the maximum of a Gaussian random field or a chi-square random process. The main results are provided in Chapter 4. Then, we discuss a simulation study under model misspecification. Chapter 5 is devoted to the analysis of growth curve data. Chapter 6 considers the statistical parametric mapping approach as future work. Details of the proofs are in the Appendix.

Chapter 2

Simultaneous Confidence Bands

Simultaneous confidence bands are useful statistical inferential tools that can be used in many statistical branches. In this chapter, we summarize simultaneous confidence bands for simple and multiple regression models.

2.1 Confidence bands for one simple regression model

It is an important task to assess where the true model $x^\top \beta$ lies in regression analysis from which the observed data have been generated.

When σ is known, a $1-\alpha$ confidence region β is given by

$$\left\{\beta : \frac{(\beta - \widehat{\beta})^\top (X^\top X)(\beta - \widehat{\beta})}{(p+1)\sigma^2} \le \chi^2_\alpha(p+1)\right\},$$

where $\chi^2_\alpha(p+1)$ is the upper α point of the χ^2 distribution with $p+1$ degrees of freedom.

When σ is unknown, a $1-\alpha$ confidence region β is given by

$$\left\{\beta : \frac{(\beta - \widehat{\beta})^\top (X^\top X)(\beta - \widehat{\beta})}{(p+1)||Y - X\widehat{\beta}||^2/(n-p-1)} \le f^\alpha_{p+1,n-p-1}\right\},$$

where $f^\alpha_{p+1,n-p-1}$ is the upper α point of the F distribution with degrees of freedom of $p+1$ and $n-p-1$.

2.2 Confidence bands for one multiple regression model

The most well-known simultaneous confidence band of level $1-\alpha$ for the regression model $x^\top\beta$ for all $x \in \mathbb{R}^p$ is given by Hotelling (1951) and Scheffé (1953, 1959). Working and Hotelling (1929) generalizes the band for a simple linear regression model. When σ is known,

$$x^\top\beta \in x^\top\widehat{\beta} \pm \sqrt{\chi^2_\alpha(p+1)}\sigma\sqrt{x^\top(X^\top X)^{-1}x}.$$

When σ is unknown,

$$x^\top\beta \in x^\top\widehat{\beta} \pm \sqrt{(p+1)f^\alpha_{p+1,n-p-1}}\widehat{\sigma}\sqrt{x^\top(X^\top X)^{-1}x}.$$

The lower parts and the upper parts of the band are symmetric about the fitted model $x^\top\widehat{\beta}$.

2.3 Confidence bands for more than two multiple regression models

Suppose k linear regression models are defined as follows

$$Y_i = X_i\beta_i + \varepsilon_i, \quad i = 1, \ldots, k,$$

where $Y_i = (y_{i,1}, \ldots, y_{i,n_i})^\top$ is a vector of random observations, X_i is a $n_i \times (p+1)$ full column-rank design matrix with the lth ($1 \le l \le n$) row given by $(1, x_{l,1}, \ldots, x_{l,p})$, $\beta_i = (\beta_{i,0}, \ldots, \beta_{i,p})^\top$, and $\varepsilon_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,n})$ with all the $\varepsilon_{i,j}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$ being independent and identically distributed (i.i.d.) $\mathcal{N}(0, \sigma^2)$ random variables. Simultaneous confidence bands for all of the contrasts between the k regression models are given as

$$\sum_{i=1}^k c_i x^\top\beta_i, \quad \text{for all} \quad c = (c_1, \ldots, c_k)^\top \in \mathcal{C},$$

where $\mathcal{C}$ is the set of all contrasts

$$\mathcal{C} = \Big\{ c = (c_1, \ldots, c_k)^\top \in \mathbb{R}^k : \sum_{i=1}^k c_i = 0 \Big\}.$$

When σ is known,

$$\sum_{i=1}^{k} c_i x^\top \beta_i \in \sum_{i=1}^{k} c_i x^\top \widehat{\beta}_i \pm c_\alpha \sigma \sqrt{\sum_{i=1}^{k} c_i^2 x^\top (X_i^\top X_i)^{-1} x}, \quad \forall x \in \mathbb{R}^{p+1}, \quad \forall c \in \mathcal{C}.$$

When σ is unknown,

$$\sum_{i=1}^{k} c_i x^\top \beta_i \in \sum_{i=1}^{k} c_i x^\top \widehat{\beta}_i \pm d_\alpha \widehat{\sigma} \sqrt{\sum_{i=1}^{k} c_i^2 x^\top (X_i^\top X_i)^{-1} x}, \quad \forall x \in \mathbb{R}^{p+1}, \quad \forall c \in \mathcal{C},$$

where c_α and d_α are determined by simulations. This provides a set of simultaneous confidence bands for all of the contrasts between the k regression models.

2.4 Comparisons of nonlinear regression curves

Considering multiple comparisons of k (≥ 3) nonlinear (curvilinear) regression curves estimated from independent k groups. Suppose that for each group $i = 1, \ldots, k$, and for each explanatory variable $x_j \in \mathcal{X}$, $j = 1, \ldots, n$, we have observations $y_{ij1}, \ldots, y_{ijr_i}$ as objective variables with r_i replications, which are assumed to follow the model

$$y_{ijh} = g_i(x_j) + \varepsilon_{ijh}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n, \quad h = 1, \ldots, r_i. \tag{2.4.1}$$

Here, $\mathcal{X} \subseteq \mathbb{R}$ is the domain of explanatory variables, and random errors ε_{ijh} are assumed to be independently distributed as the normal distribution $\mathcal{N}(0, \sigma(x_j)^2)$. The variance function $\sigma(x)^2$ is supposed to be known, or at least known up to a constant $\sigma(x)^2 = \sigma^2 \sigma_0(x)^2$. In the case of the latter, we suppose that an independent estimator $\widehat{\sigma}^2$ of σ^2 is available. In addition, we assume that the true regression curve has the form

$$g_i(x) = \beta_i^\top f(x), \quad x \in \mathcal{X}, \tag{2.4.2}$$

where $f(x) = (f_1(x), \ldots, f_p(x))^\top$ is a known regression basis vector function, and $\beta_i = (\beta_{i1}, \ldots, \beta_{ip})^\top$ is an unknown parameter vector. Then, the least squares estimator $\widehat{\beta}_i$ of β_i has the multivariate normal distribution $\mathcal{N}_p(\beta_i, r_i^{-1}\Sigma)$, where

$$\Sigma = \left(\sum_{j=1}^{n} \frac{1}{\sigma(x_j)^2} f(x_j) f(x_j)^\top \right)^{-1}$$

is the inverse of the $p \times p$ information matrix. When $\sigma(x)^2 = \sigma^2 \sigma_0(x)^2$, we have $\Sigma = \sigma^2 \Sigma_0$, where Σ_0 is Σ with $\sigma(x_j)$ replaced by $\sigma_0(x_j)$.

Let $\mathcal{C}$ denote the set of vectors $c = (c_1, \ldots, c_k)^\top$ such that $\sum_{i=1}^k c_i = 0$. The focus of this thesis is the construction of $1-\alpha$ simultaneous confidence bands for all the contrasts $\sum_{i=1}^k c_i g_i(x) = \sum_{i=1}^k c_i \beta_i^\top f(x)$ between the k regression curves for all $x \in \mathcal{X}$ and $c \in \mathcal{C}$, where $\mathcal{X}$ is a given finite interval $[a, b]$, a finite union of intervals $\bigsqcup_i [a_i, b_i]$, or an infinite interval $(-\infty, \infty)$, with the symbol '$\bigsqcup$' denoting a disjoint union.

Specifically, according to the traditional form of the point estimate plus or minus a probability point times the estimated standard error, we construct a $1-\alpha$ simultaneous confidence band of the form

$$\sum_{i=1}^k c_i \beta_i^\top f(x) \in \sum_{i=1}^k c_i \widehat{\beta}_i^\top f(x) \pm b_{1-\alpha} \sqrt{\left(\sum_{i=1}^k \frac{c_i^2}{r_i}\right) f(x)^\top \Sigma f(x)}, \tag{2.4.3}$$

where $\widehat{\beta}_i^\top f(x)$ is the estimator of $\beta_i^\top f(x)$ in (2.4.2). This form is referred to as a hyperbolic-type (Liu, 2010). The critical value $b_{1-\alpha}$ is determined such that the event in (2.4.3) for all $x \in \mathcal{X}$ and $c \in \mathcal{C}$ holds with a probability of at least $1-\alpha$. Our problem typically arises from growth curve analysis and longitudinal data analysis.

Throughout this paper, we assume that the regression curve $g_i(x)$ is a linear combination of a finite number of known basis functions in (2.4.2). Although it is a conventional regression model, we must always be careful regarding the approximation bias caused by model misspecification. This issue is examined in Section 4.4.

2.5 The problem we considered

The problem concerning the construction of simultaneous confidence bands in a regression model originates with Working and Hotelling (1929). They formalized this problem as the construction of confidence intervals for an estimated regression line, and provided a critical value by making use of the Cauchy-Schwarz inequality. Specifically, Working and Hotelling (1929) treated the case of

(i) one regression model (equivalent to case $k = 2$ in our problem),
(ii) the simple regression $f(x) = (1, x)^\top$, and
(iii) the unrestricted domain of the explanatory variables $\mathcal{X} = (-\infty, \infty)$.

Subsequently, many reports concerning the relaxation of these conditions have appeared in literature.

In the case of one regression model, Wynn and Bloomfield (1971) pointed out that the use of the Cauchy-Schwarz inequality leads to conservative bands unless both (ii) and (iii) hold. They illustrated improved confidence bands for the quadratic regression $f(x) = (1, x, x^2)^\top$. Uusipaikka (1983) constructed exact confidence bands for linear regression when $\mathcal{X}$ is a finite interval. See Liu, Lin, and Piegorsch (2008) and Liu

(2010) for historical reviews. The problem of $k \geq 3$ regression curve comparisons was considered by Spurrier (1999, 2002) and Lu and Chen (2009), who proposed procedures based on simple linear regression. However, it is difficult to extend these methods to nonlinear regression.

One exception is Naiman (1986)'s integral-geometric approach. In the unit sphere $\mathbb{S}^{p-1}$ of the p-dimensional Euclidean space, he defined a trajectory

$$\Gamma = \overline{\{\psi(x) \mid x \in \mathcal{X}\}} \subset \mathbb{S}^{p-1} \tag{2.5.1}$$

of a normalized basis vector function

$$\psi(x) = \frac{\Sigma^{1/2} f(x)}{\|\Sigma^{1/2} f(x)\|}, \tag{2.5.2}$$

and evaluated the volume of the Γ tubular neighborhood. In the case of one regression model, he constructed a simultaneous confidence band with the critical value obtained from this volume. The volume formula for such tubes originated from Hotelling (1939) and Weyl (1939). Currently, this idea is understood in the volume-of-tube method framework (Adler and Taylor (2007), Kuriki and Takemura (2001), Kuriki and Takemura (2009), Sun (1993), Takemura and Kuriki (2002)). As shown in Section 4.1, we require the tail probability of the maximum of a Gaussian random field or chi-square random process as a pivotal quantity. Volume-of-tube is a methodology to evaluate such tail probabilities.

In this paper, we adopt this integral-geometric approach. In the case of $k \geq 3$, we define a subset M in (4.2.1) of a unit sphere, and by evaluating the volume of its tubular neighborhood, we obtain the critical value $b_{1-\alpha}$ in (2.4.3) by means of the volume-of-tube method. Moreover, we prove that the proposed confidence band is conservative. It is known that Naiman (1986)'s confidence band is conservative (Naiman's inequality, see also Johnstone and Siegmund (1989)), and our result is regarded as its generalization.

Note that, in the setting of this paper, the covariance matrices of the estimators $\widehat{\beta}_i$ are identical up to a multiplicative constant. This property arises from the condition that the explanatory variables x_j are common between k groups in the model (2.4.1). This represents the purported balanced case. For the unbalanced case, the problem of constructing simultaneous confidence bands is quite tedious and only simulation-based approaches are available (Jamshidian, Liu, and Bretz (2010), Liu (2010), Liu, Jamshidian, and Zhang (2004), Liu, Wynn, and Hayter (2008)). In this paper, we address only the balanced case.

Moreover, note that in the one-group case ($k = 1$), various simultaneous confidence bands by means of the volume-of-tube method have been proposed. Johansen and Johnstone (1990) demonstrated the usefulness of Hotelling's volume formula for the construction of simultaneous bands. The application to the B-spline regression is found in Shen, Wolfe, and Zhou (1998). Sun and Loader (1994) proposed a modification to the volume-of-tube formula when a small approximation bias caused by model misspecification exists. In succeeding papers, Sun and her coauthors developed this idea further in various model settings (Faraway and Sun (1995), Sun, Loader, and McCormick (2000), Sun, Raz, and Faraway (1999)). See also Krivobokova, Kneib, and Claeskenset (2010). The crucial difference between this paper and existing work is that in this paper, we need to treat a Gaussian random field with a general dimensional ($k-1$ dimensional) index set, and need the volume formula up to an arbitrary order.

Chapter 3

The Volume-of-Tube Method

3.1 Definition of the tube

Considering general case, let $\mathbb{S}^{n-1} = \mathbb{S}(\mathbb{R}^n)$ be the unit sphere in $\mathbb{R}^n$ and let $M \subset \mathbb{S}^{n-1}$ be a closed subset of $\mathbb{S}^{n-1}$. Let the elements of $\xi = (\xi_1, \ldots, \xi_n)$ be independent and standard normal random variables. (We write this as $\xi \sim \mathcal{N}_n(0, I_n)$.) $\langle \cdot, \cdot \rangle$ denotes the standard inner product of $\mathbb{R}^n$. Our problem is to find the distribution of the maximum of the Gaussian random field $X(p) = \langle \xi, p \rangle$, $p \in M$:

$$\Pr\Big(\max_{p \in M} \langle \xi, p \rangle \geq c\Big). \tag{3.1.1}$$

Definition 3.1.1 (Tube)**.** *The tube (spherical tube) of radius θ about M is defined to be the set of points on $\mathbb{S}^{n-1}$ whose great-circle distance to M is less than or equal to θ :*

$$M_\theta = \Big\{ q \in \mathbb{S}^{n-1} \mid \mathrm{dist}(q, M) \leq \theta \Big\} = \Big\{ v \in \mathbb{S}^{n-1} \mid \min_{u \in M} \cos^{-1}\big(u^\top v\big) \leq \theta \Big\}.$$

For an n-dimensional standard normal random vector $\xi \sim \mathcal{N}_n(0, I_n)$, its "length" $\|\xi\|$ and its "direction" $\zeta = \xi / \|\xi\|$ are independently distributed and the distribution of ζ is the uniform distribution over the unit sphere $\mathrm{Unif}(\mathbb{S}^{n-1})$. Hence,

$$\begin{aligned}
\Pr\Big(\max_{p \in M} \langle \xi, p \rangle \geq c\Big) =& \mathrm{E}\left[\Pr\Big(\max_{p \in M} \langle \zeta, p \rangle \geq \frac{c}{\|\xi\|} \mid \|\xi\| \Big) \right] \\
=& \mathrm{E}\left[\Pr\Big(\mathrm{dist}(\zeta, M) \leq \cos^{-1}\Big(\frac{c}{\|\xi\|}\Big) \mid \|\xi\| \Big) \right] \\
=& \frac{1}{\mathrm{Vol}(\mathbb{S}^{n-1})} \mathrm{E}\left[\mathrm{Vol}\Big(M_{\cos^{-1}(c/\|\xi\|)} \Big) \right],
\end{aligned}$$

where $\mathrm{Vol}(\cdot)$ is the $(n-1)$-dimensional volume. If the volume of the tube $\mathrm{Vol}(M_\theta)$ can be evaluated for every θ, then we can integrate it once (that is, we can take the expected value with respect to $\|\xi\|$) to obtain the tail probability of the maximum (3.1.1).

3.2 Definition of the critical radius

The support cone (or tangent cone) of M at $u \in M$ is denoted by $S_u M$. (See Section 1.2 of Takemura and Kuriki (2002) for the definition.) The cone with base set M is denoted by

$$\mathrm{co}(M) = \bigsqcup_{\lambda \geq 0} \lambda M.$$

Then, the support cone of $\mathrm{co}(M)$ at $u \in M$ is decomposed as $S_u(\mathrm{co}(M)) = S_u M \oplus \mathrm{span}\{u\}$, where $\mathrm{span}\{u\}$ is the linear space spanned by u. The normal cone of $\mathrm{co}(M)$ at $u \in M$ is defined by the dual of the support cone: $N_u(\mathrm{co}(M)) = S_u(\mathrm{co}(M))^*$.

Definition 3.2.1 (Critical radius)**.** *We say that the tube M_θ does not have a self-intersection if every point $q \in M_\theta \setminus M$ is uniquely written as*

$$q = p\cos\psi + v\sin\psi, \qquad p \in M,\ v \in \ N_u(\mathrm{co}(M)) \cap \mathbb{S}^{n-1},\ \psi \in (0,\theta].$$

The supremum of the radius θ such that M_θ does not have a self-intersection

$$\theta_c = \sup\{\theta \geq 0 \mid M_\theta \text{ does not have a self-intersection}\}$$

is the critical radius (reach) of M (Figure 3.2.1). Let $\theta_c = \pi/2$ when θ_c is more than $\pi/2$.

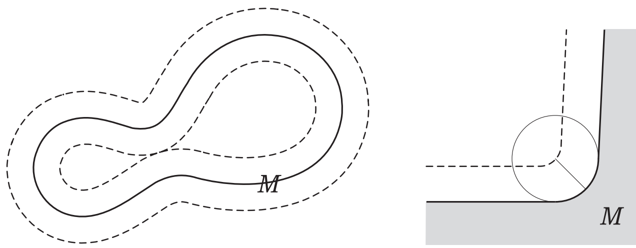

Figure 3.2.1: Tubes with a radius equal to the critical radius (Kuriki and Takemura, 2009).

3.3 Volume-of-tube method and upper tail probability

In this section, we summarize the volume-of-tube method for evaluating the upper tail probability of the maximum of a Gaussian random field.

Let ξ be a Gaussian random vector distributed as $\mathcal{N}_n(0, I)$. Let M be a closed subset of $\mathbb{S}^{n-1}$, the unit sphere (the set of unit column vectors) of $\mathbb{R}^n$. Then, the random map $u \mapsto \xi^\top u$, $u \in M$, is a Gaussian random field with mean 0, variance 1, and a covariance function $\mathrm{Cov}(\xi^\top u, \xi^\top v) = u^\top v$. The volume-of-tube method approximates the distribution of the maximum $\max_{u\in M} \xi^\top u$. To apply the volume-of-tube method, we require the following assumption on M.

Assumption 3.3.1. *M is a d-dimensional closed piecewise C^2-manifold, or M is a d-dimensional C^2-manifold with piecewise C^2-boundary. We write $M = \mathrm{Int}M \sqcup \partial M$, where $\mathrm{Int}M$ and ∂M denote the interior and the boundary of M, respectively. In the former case, $\partial M = \emptyset$.*

Under Assumption 3.3.1, we can prove that $\theta_c > 0$.

Note that the $(m-1)$-dimensional volume of $\mathbb{S}^{m-1}$ is $\Omega_m = 2\pi^{m/2}/\Gamma(m/2)$. For $m \times m$ matrix $A = (a_{ij})$, let $\mathrm{tr}_0 A = 1$ and

$$\mathrm{tr}_e A = \sum_{1\le k_1 < \ldots < k_e \le m} \det(a_{k_i k_j})_{1\le i,j\le e}, \quad 1 \le e \le m$$

(Muirhead (2005), Appendix A.7). Note that $\mathrm{tr}_1 A = \mathrm{tr}A$, $\mathrm{tr}_m A = \det A$. The upper probability of the chi-square distribution with m degrees of freedom is denoted by $\overline{G}_m(\cdot)$. Now we can provide the upper tail probability formula for the Gaussian field $\xi^\top u$, $u \in M$. The theorem below is a special case of Proposition 2.2 of Takemura and Kuriki (2002).

Proposition 3.3.1. *As $b \to \infty$,*

$$\Pr\left(\max_{u\in M} \xi^\top u \ge b\right) = \overline{P}_{\text{tube}}(b) + O(\overline{G}_n(b^2(1+\tan^2\theta_{\text{c}}))), \tag{3.3.1}$$

where

$$\overline{P}_{\text{tube}}(b) = \sum_{0\le e\le d,\, e:\text{even}} w_{d+1-e}\overline{G}_{d+1-e}(b^2) + \sum_{0\le e\le d-1} w'_{d-e}\overline{G}_{d-e}(b^2), \tag{3.3.2}$$

with

$$w_{d+1-e} = \frac{1}{\Omega_{d+1-e}\Omega_{n-d-1+e}} \int_{\text{Int}M} \left\{ \int_{N_u(\text{co}(M))\cap\mathbb{S}^{n-1}} \text{tr}_e H(u,v)\,\text{d}v \right\} \text{d}u, \tag{3.3.3}$$

$$w'_{d-e} = \frac{1}{\Omega_{d-e}\Omega_{n-d+e}} \int_{\partial M} \left\{ \int_{N_u(\text{co}(M))\cap\mathbb{S}^{n-1}} \text{tr}_e H'(u,v)\,\text{d}v \right\} \text{d}u. \tag{3.3.4}$$

Here, $H(u,v)$ is the second fundamental form of $\text{Int}M$ at u in the direction of v, and $H'(u,v)$ is the second fundamental form of ∂M at u in the direction of v. du is the volume element of $\text{Int}M$ or ∂M, and dv is the volume element of $N_u(\text{co}(M))\cap\mathbb{S}^{n-1}$.

In (3.3.1), because $\theta_{\text{c}} > 0$, the error term $O\big(\overline{G}_n(b^2(1+\tan^2\theta_{\text{c}}))\big) = O(b^{n-2}e^{-b^2(1+\tan^2\theta_{\text{c}})/2})$ is exponentially smaller than each term $\overline{G}_j(b^2) = O(b^{j-2}e^{-b^2/2})$. Hence, (3.3.2) can be used as an approximation formula when b is large. The method in which $\overline{P}_{\text{tube}}(b)$ is used as an approximate value is referred to as the volume-of-tube method, or simply, the tube method. This name comes from the volume formula for M_θ below.

Remark 3.3.1. *For the radius $\theta \in [0, \theta_{\text{c}}]$, the $(n-1)$-dimensional spherical volume of the tube M_θ is given by*

$$\text{Vol}_{n-1}(M_\theta) = \Omega_n \Bigg\{ \sum_{0\le e\le d,\, e:\text{even}} w_{d+1-e}\overline{B}_{\frac{1}{2}(d+1-e),\frac{1}{2}(n-d-1+e)}(\cos^2\theta) + \sum_{0\le e\le d-1} w'_{d-e}\overline{B}_{\frac{1}{2}(d-e),\frac{1}{2}(n-d+e)}(\cos^2\theta) \Bigg\},$$

where w_{d+1-e} and w'_{d-e} are given in (3.3.3) and (3.3.4), $\overline{B}_{a,b}(\cdot)$ is the upper probability of the beta distribution with parameter (a,b).

The critical radius θ_{c} can be evaluated using the following characterization (Theorem 4.18 of Federer (1959), Proposition 4.3 of Johansen and Johnstone (1990), Lemma 2.2 of Takemura and Kuriki (2002)). For a proof, see Theorem 2.9 of Kuriki and Takemura (2009).

Proposition 3.3.2. *The critical radius* θ_{c} *of* M *is given by*

$$\tan^2\theta_{\mathrm{c}} = \inf_{u\neq v\in M} \frac{(1-u^\top v)^2}{\|P_v^\perp(u-v)\|^2}, \tag{3.3.5}$$

where $P_v^\perp$ *is the orthogonal projection onto the normal cone* $N_v(\mathrm{co}(M))$ *of* $\mathrm{co}(M)$ *at* v.

The local critical radius $\theta_{\mathrm{c,loc}}$ is defined as

$$\tan^2\theta_{\mathrm{c,loc}} = \liminf_{u\neq v\in M,\, \|u-v\|\to 0} \frac{(1-u^\top v)^2}{\|P_v^\perp(u-v)\|^2}. \tag{3.3.6}$$

From the definition, it holds that $\theta_{\mathrm{c}} \le \theta_{\mathrm{c,loc}}$. In general, $\theta_{\mathrm{c,loc}}$ is easier to evaluate than θ_{c}.

3.4 Expected Euler-characteristic heuristic

We have summarized the volume-of-tube method to evaluate the upper tail probabilities of the maximum of random fields thus far. There is another method utilized for the same purpose, known as the expected Euler-characteristic heuristic (Adler and Taylor (2007), Worsley (1995)). When applied to the Gaussian random field $\xi^\top u$, $u \in M$, this method is stated as follows: For each b, define the excursion set by

$$A_b = \{u \in M \mid \xi^\top u \ge b\}.$$

Let $\chi(\cdot)$ be the Euler-Poincaré characteristic of a set, and $\mathbb{1}(\cdot)$ be the indicator function for an event. The expected Euler-characteristic heuristic assumes that $\mathbb{1}(A_b \neq \emptyset) \approx \chi(A_b)$ for large b, and

$$\Pr\left(\max_{u\in M} \xi^\top u \ge b\right) = \mathrm{E}\{\mathbb{1}(A_b \neq \emptyset)\} \approx \mathrm{E}\{\chi(A_b)\}.$$

Note that $\chi(A_b)$ can be evaluated by Morse's theorem, and is more tractable than $\mathbb{1}(A_b \neq \emptyset)$. Takemura and Kuriki (2002) proved the equivalence of the volume-of-tube method and expected Euler-characteristic heuristic as follows.

Proposition 3.4.1 (Proposition 3.3 of Takemura and Kuriki (2002))**.**

$$\mathrm{E}\{\chi(A_b)\} = \overline{P}_{\mathrm{tube}}(b), \quad \textit{for all } b \ge 0.$$

Using this, Takemura and Kuriki (2002) provided an alternative proof that the confidence band of Naiman (1986) is conservative.

Chapter 4

Construction of Simultaneous Confidence Bands

4.1 Random fields as pivotal quantities

Our problem is to determine the critical value $b_{1-\alpha}$ in (2.4.3). First, assume that Σ is fully known. Define a pivotal quantity:

$$T(x,c) = \frac{\sum_{i=1}^k c_i(\widehat{\beta}_i - \beta_i)^\top f(x)}{\sqrt{\left(\sum_{i=1}^k \frac{c_i^2}{r_i}\right) f(x)^\top \Sigma f(x)}}. \tag{4.1.1}$$

Then, the critical value $b_{1-\alpha}$ is solution b of the equation:

$$\Pr\bigl\{T(x,c) \le b,\ \forall x \in \mathcal{X},\ \forall c \in \mathcal{C}\bigr\} = \Pr\left\{\max_{x\in\mathcal{X},c\in\mathcal{C}} T(x,c) \le b\right\} = 1-\alpha.$$

In this expression, we use $T(x,c)$ instead of $|T(x,c)|$, because $c \in \mathcal{C}$ implies $-c \in \mathcal{C}$ and $|T(x,c)|$ is equal to $T(x,c)$ or $T(x,-c)$. Inverting $|T(c,x)| \le b_{1-\alpha}$ yields the $1-\alpha$ simultaneous confidence band in (2.4.3).

In the following, we show that $b_{1-\alpha}^2$ is the upper α point of the maximum of a chi-square random process. We can assume that $\sum_{i=1}^k c_i^2/r_i = 1$ without the loss of generality, because $T(x,c)$ is a homogeneous function in c. Let $\rho = (\sqrt{r_1},\dots,\sqrt{r_k})^\top$, and define a $k\times(k-1)$ matrix H such that $\rho^\top H = 0$, $H^\top H = I_{k-1}$, and $HH^\top = I_k - \rho\rho^\top/(\rho^\top\rho)$. (An example of H is given in Remark 4.2.1 below.) Then, the $c = (c_1,\dots,c_k)^\top$ such that $\sum_{i=1}^k c_i^2/r_i = 1$ and $\sum_{i=1}^k c_i = 0$ are represented as

$$c = \mathrm{diag}(\sqrt{r_1},\dots,\sqrt{r_k})Hh, \quad h \in \mathbb{S}^{k-2},$$

where $\mathbb{S}^{k-2}$ is the set of $(k-1)$-dimensional unit column vectors.

Let $\Sigma^{1/2}$ be a matrix such that $(\Sigma^{1/2})^\top \Sigma^{1/2} = \Sigma$, and let $\Sigma^{-1/2}$ be its inverse. Then, $\eta_i = \sqrt{r_i}(\Sigma^{-1/2})^\top(\widehat{\beta}_i - \beta_i)$ is distributed normally as $\mathcal{N}_p(0, I)$, independently for $i = 1, \ldots, k$. Let $\psi : \mathcal{X} \to \mathbb{S}^{p-1}$ as defined in (2.5.2). Then, $T(x, c)$ is rewritten as

$$\begin{aligned}
T(x,c) =& \sum_{i=1}^{k} \frac{c_i}{\sqrt{r_i}} \sqrt{r_i}\{(\Sigma^{-1/2})^\top(\widehat{\beta}_i - \beta_i)\}^\top \frac{\Sigma^{1/2} f(x)}{\|\Sigma^{1/2} f(x)\|} \\
=& c^\top \mathrm{diag}(\sqrt{r_1}, \ldots, \sqrt{r_k})^{-1} \begin{pmatrix} \eta_1^\top \\ \vdots \\ \eta_k^\top \end{pmatrix}_{k\times p} \psi(x) \\
=& h^\top \begin{pmatrix} \xi_1^\top \\ \vdots \\ \xi_{k-1}^\top \end{pmatrix}_{(k-1)\times p} \psi(x) \\
=& \xi^\top \{h \otimes \psi(x)\},
\end{aligned} \tag{4.1.2}$$

where ξ_i are $p \times 1$ vectors defined by $(\xi_1, \ldots, \xi_{k-1})_{p\times(k-1)} = (\eta_1, \ldots, \eta_k)_{p\times k} H$, $\xi = (\xi_1^\top, \ldots, \xi_{k-1}^\top)^\top$ is a $p(k-1) \times 1$ vector, and '$\otimes$' is the Kronecker product. Vectors η_i consist of independent standard Gaussian random variables $\mathcal{N}(0, 1)$, therefore, so does vector ξ. When x and h are fixed, because $\|\psi(x)\| = \|h \otimes \psi(x)\| = 1$, $\xi_i^\top \psi(x)$ is distributed as $\mathcal{N}(0, 1)$ independently for $i = 1, \ldots, k$, and $\xi^\top\{h \otimes \psi(x)\}$ is distributed as $\mathcal{N}(0, 1)$.

From (4.1.2), we can see that

$$\max_{c\in\mathcal{C}} T(x, c) = \sqrt{\sum_{i=1}^{k-1} \left\{\xi_i^\top \psi(x)\right\}^2}. \tag{4.1.3}$$

For each fixed x, this is distributed as the square root of the chi-square distribution χ^2_{k-1} with $k - 1$ degrees of freedom.

When $\Sigma = \sigma^2 \Sigma_0$ with Σ_0 known, and an independent estimator $\widehat{\sigma}^2 \sim \sigma^2 \chi^2_\nu / \nu$ of unknown σ^2 is available, we redefine $T(x, c)$ in (4.1.1) by replacing Σ in the denominator with $\widehat{\sigma}^2 \Sigma_0$. Thus, instead of (4.1.2) and (4.1.3) we have

$$T(x, c) = \frac{1}{\tau} \xi^\top \{h \otimes \psi(x)\}, \quad \max_{c\in\mathcal{C}} T(x, c) = \sqrt{\frac{1}{\tau^2} \sum_{i=1}^{k-1} \left\{\xi_i^\top \psi(x)\right\}^2}, \quad \tau^2 = \frac{\widehat{\sigma}^2}{\sigma^2}.$$

4.2 Tube formula

In the particular case of the problem we consider, the maximum of $Z(x,h)$ in (4.2.2) can be treated in this framework by setting

$$M = \overline{\{h \otimes \psi(x) \mid (x,h) \in \mathcal{X} \times \mathbb{S}^{k-2}\}} \quad \text{and} \quad n = p(k-1). \tag{4.2.1}$$

The dimension of M is $d = \dim M = k-1$.

When M is defined by (4.2.1), we can provide a sufficient condition for Assumption 3.3.1.

Assumption 4.2.1. *$\psi : \mathcal{X} \to \mathbb{S}^{p-1}$ is a one-to-one map of class piecewise C^2. There does not exist $x, \tilde{x} \in \mathcal{X}$ such that $\psi(x) = -\psi(\tilde{x})$.*

Under Assumption 4.2.1, the map $(x,h) \mapsto h \otimes \psi(x)$ is a piecewise C^2 one-to-one map.

Example 4.2.1. *Consider the polynomial regression with a basis function vector $f(x) = (1, x, \ldots, x^{p-1})^\top$. When the domain of x is a finite interval $\mathcal{X} = [a,b]$, we have*

$$\begin{aligned} \mathrm{Int} M =& \{h \otimes \psi(x) \mid x \in (a,b),\, h \in \mathbb{S}^{k-2}\}, \\ \partial M =& \{h \otimes \psi(a) \mid h \in \mathbb{S}^{k-2}\} \sqcup \{h \otimes \psi(b) \mid h \in \mathbb{S}^{k-2}\}. \end{aligned}$$

When $\mathcal{X} = (-\infty, \infty)$, $\psi(\pm\infty) = (\pm 1)^{p-1} \Sigma^{1/2} e_p / \sqrt{e_p^\top \Sigma e_p}$ with $e_p = (0, \ldots, 0, 1)^\top$, and hence $h \otimes \psi(\infty) = (-1)^{p-1} h \otimes \psi(-\infty)$. This denotes that M is a closed manifold without boundary.

Example 4.2.2. *Consider the trigonometric regression with a basis function vector*

$$f(x) = \Big(1, \sqrt{2}\cos x, \sqrt{2}\sin x, \ldots, \sqrt{2}\cos mx, \sqrt{2}\sin mx\Big)^\top.$$

When $\mathcal{X} = [0, 2\pi)$, M is a closed manifold without boundary.

Now, we consider the object in (4.1.2) as a random function of (x,h):

$$Z(x,h) = \xi^\top \{h \otimes \psi(x)\}, \quad (x,h) \in \mathcal{X} \times \mathbb{S}^{k-2}, \tag{4.2.2}$$

where $\xi \sim \mathcal{N}_{p(k-1)}(0, I)$. Then, $Z(x,h)$ is the Gaussian random field with mean 0, variance 1, and covariance function

$$\mathrm{Cov}\big[Z(x,h), Z(\tilde{x}, \tilde{h})\big] = \psi(x)^\top \psi(\tilde{x}) \cdot h^\top \tilde{h}.$$

Similarly, we define the chi-square random process with $k-1$ degrees of freedom:

$$Y(x) = \sum_{i=1}^{k-1} \{\xi_i^\top \psi(x)\}^2, \quad x \in \mathcal{X}. \tag{4.2.3}$$

We summarize the results of this section below.

Theorem 4.2.1. *When Σ is known, the critical value $b_{1-\alpha}$ is determined as the solution $b = b_{1-\alpha}$ of*

$$\Pr\Big\{ \max_{x\in\mathcal{X}, h\in\mathbb{S}^{k-2}} Z(x,h) \ge b \Big\} = \Pr\Big\{ \max_{x\in\mathcal{X}} Y(x) \ge b^2 \Big\} = \alpha,$$

where $Z(x,h)$ is the Gaussian random field defined in (4.2.2), and $Y(x)$ is the chi-square random process defined in (4.2.3).

When $\Sigma = \sigma^2\Sigma_0$ with Σ_0 known, the critical value $b_{1-\alpha}$ is determined as the solution $b = b_{1-\alpha}$ of

$$\mathrm{E}\Big[\Pr\Big\{ \max_{x\in\mathcal{X}, h\in\mathbb{S}^{k-2}} Z(x,h) \ge b\tau \,\Big|\, \tau^2 \Big\}\Big] = \mathrm{E}\Big[\Pr\Big\{ \max_{x\in\mathcal{X}} Y(x) \ge b^2\tau^2 \,\Big|\, \tau^2 \Big\}\Big] = \alpha,$$

where the expectation is taken over $\tau^2 \sim \chi^2_\nu/\nu$, with ν being the degrees of freedom of the estimator of σ^2.

Remark 4.2.1. *An example of $k\times(k-1)$ matrix H such that $\rho^\top H = 0$, $H^\top H = I_{k-1}$, $HH^\top = I_k - \rho\rho^\top/(\rho^\top\rho)$ with $\rho = (\sqrt{r_1},\ldots,\sqrt{r_k})^\top$ is given as*

$$H = \begin{pmatrix} \frac{\sqrt{r_1 r_2}}{\sqrt{R_1 R_2}} & \frac{\sqrt{r_1 r_3}}{\sqrt{R_2 R_3}} & \cdots & \frac{\sqrt{r_1 r_k}}{\sqrt{R_{k-1} R_k}} \\ -\frac{R_1}{\sqrt{R_1 R_2}} & \frac{\sqrt{r_2 r_3}}{\sqrt{R_2 R_3}} & \cdots & \frac{\sqrt{r_2 r_k}}{\sqrt{R_{k-1} R_k}} \\ & -\frac{R_2}{\sqrt{R_2 R_3}} & \cdots & \frac{\sqrt{r_3 r_k}}{\sqrt{R_{k-1} R_k}} \\ & & \ddots & \vdots \\ 0 & & & -\frac{R_{k-1}}{\sqrt{R_{k-1} R_k}} \end{pmatrix}_{k\times(k-1)},$$

where $R_i = \sum_{j=1}^i r_j$.

Theorem 4.2.2. *Let $\xi \sim \mathcal{N}_n(0, I)$, $n = p(k-1)$. Let $\Gamma \subset \mathbb{S}^{p-1}$ and $M \subset \mathbb{S}^{n-1}$ be defined by (2.5.1) and (4.2.1), and let $|\Gamma|$ denote the length of Γ. Assume Assumption*

4.2.1 on ψ. Then, as $b \to \infty$,

$$\begin{aligned}\Pr\Big\{\max_{(x,h)\in\mathcal{X}\times\mathbb{S}^{k-2}} Z(x,h) \ge b\Big\} &= \Pr\Big\{\max_{x\in\mathcal{X}} Y(x) \ge b^2\Big\} \\ &= \Pr\Big(\max_{u\in M} \xi^\top u \ge b\Big) \\ &= \overline{P}_{\rm tube}(b) + O\big(b^{n-2}e^{-(1+\tan^2\theta_{\rm c})b^2/2}\big),\end{aligned}$$

where

$$\overline{P}_{\rm tube}(b) = \frac{\Gamma(\frac{k}{2})}{\sqrt{\pi}\,\Gamma(\frac{k-1}{2})}|\Gamma|\big\{\overline{G}_k(b^2) - \overline{G}_{k-2}(b^2)\big\} + \chi(\Gamma)\overline{G}_{k-1}(b^2). \tag{4.2.4}$$

Note that if Γ (and hence M) has no boundary, then Γ is homeomorphic to $\mathbb{S}^1$, and therefore $\chi(\Gamma) = 0$. Otherwise, $\chi(\Gamma)$ is the number of connected components of Γ.

Theorem 4.2.3. *Assume Assumption 4.2.1. Suppose that Γ has boundaries. The approximation formula given in Theorem 4.2.2 is a conservative bound, specifically,*

$$\Pr\Big(\max_{u\in M} \xi^\top u \ge b\Big) \le \overline{P}_{\rm tube}(b) \quad \text{for all } b \ge 0.$$

Proof. Arrange the $p(k-1)\times 1$ vector $\xi = (\xi_1^\top, \ldots, \xi_{k-1}^\top)^\top$, and define a $(k-1)\times p$ matrix $\Xi = (\xi_1, \ldots, \xi_{k-1})^\top$. Let

$$\begin{aligned} A_b &= \{u \in M \mid \xi^\top u \ge b\} = \{h \otimes q \mid (q,h) \in \Gamma \times \mathbb{S}^{k-2},\ h^\top \Xi q \ge b\} \subset \mathbb{S}^{p(k-1)-1}, \\ \widetilde{A}_b &= \{(q,h) \in \Gamma \times \mathbb{S}^{k-2} \mid h^\top \Xi q \ge b\} \subset \mathbb{S}^{p-1} \times \mathbb{S}^{k-2}, \\ B_b &= \{q \in \Gamma \mid q^\top \Xi^\top \Xi q \ge b^2\} \subset \mathbb{S}^{p-1}. \end{aligned}$$

Note that A_b is the excursion set of the Gaussian random field $\xi^\top u$, $u \in M$, $\widetilde{A}_b$ is the excursion set of the Gaussian random field $\sum_{i=1}^{k-1} h_i(\xi_i^\top q) = h^\top \Xi q$, $(q,h) \in \Gamma \times \mathbb{S}^{k-2}$, and B_b is the excursion set of the chi-square random process $\sum_{i=1}^{k-1}(\xi_i^\top q)^2 = q^\top \Xi^\top \Xi q$, $q \in \Gamma$. We will prove that for each fixed ξ, $\mathbb{1}(A_b \ne \emptyset) = \mathbb{1}(\widetilde{A}_b \ne \emptyset) = \mathbb{1}(B_b \ne \emptyset)$ and $\chi(A_b) = \chi(\widetilde{A}_b) = \chi(B_b)$.

First, note that owing to Assumption 4.2.1, the map $(q,h) \mapsto h \otimes q$ is one-to-one. Hence, A_b and $\widetilde{A}_b$ are homeomorphic and therefore $\mathbb{1}(A_b \ne \emptyset) = \mathbb{1}(\widetilde{A}_b \ne \emptyset)$ and $\chi(A_b) = \chi(\widetilde{A}_b)$.

Moreover, noting that $\widetilde{A}_b \ne \emptyset \Leftrightarrow \max_h h^\top \Xi q \ge b$ for some $q \Leftrightarrow q^\top \Xi\Xi^\top q \ge b^2$ for some $q \Leftrightarrow B_b \neq \emptyset$, that is, $\mathbb{1}(\widetilde{A}_b \neq \emptyset) = \mathbb{1}(B_b \neq \emptyset)$, we can write

$$\widetilde{A}_b = \bigsqcup_{q \in B_b} \{(q,h) \mid h \in \mathbb{S}^{k-2},\ h^\top \Xi q \geq b\}.$$

Given $b \geq 0$, the set $\{h \in \mathbb{S}^{k-2} \mid h^\top \Xi q \geq b\}$ is contractible and star-shaped about the point $h^*(q) = \Xi q / \|\Xi q\|$. That is, the map

$$\varphi : \widetilde{A}_b \times [0,1] \to \widetilde{A}_b, \quad (q,h,t) \mapsto \left(q, \frac{(1-t)h + th^*(q)}{\|(1-t)h + th^*(q)\|}\right)$$

is continuous, and $\varphi\big(\widetilde{A}_b \times \{0\}\big) = \widetilde{A}_b$ is homotopy equivalent to the set $\varphi\big(\widetilde{A}_b \times \{1\}\big) = \bigsqcup_{q \in B_b} \{(q, h^*(q))\}$. This is homotopy equivalent to $\bigsqcup_{q \in B_b} \{q\} = B_b$. Hence, $\chi(\widetilde{A}_b) = \chi(B_b)$.

Recall that B_b is the excursion set of the chi-square random process on the one-dimensional index set Γ. This means that B_b is also one-dimensional, and $\chi(B_b)$ is only the number of connected components of B_b. Therefore $\mathbb{1}(B_b \neq \emptyset) \leq \chi(B_b)$. By taking expectations,

$$\begin{aligned}\Pr\left(\max_{u \in M} \xi^\top u \geq b\right) &= \mathrm{E}\{\mathbb{1}(A_b \neq \emptyset)\} = \mathrm{E}\{\mathbb{1}(B_b \neq \emptyset)\} \\ &\leq \mathrm{E}\{\chi(B_b)\} = \mathrm{E}\{\chi(A_b)\} = \overline{P}_{\mathrm{tube}}(b).\end{aligned}$$

The last equality is owing to Proposition 3.4.1. □

Remark 4.2.2. *Naiman (1986) proved that application of the volume-of-tube method to a Gaussian random process with a one-dimensional index set always provides a conservative band. Theorem 4.2.3 is a generalization of Naiman (1986)'s inequality to a chi-square random process.*

Theorem 4.2.4. *The interior and boundary of* Γ *are denoted by* $\mathrm{Int}\Gamma$ *and* $\partial\Gamma$*, respectively. The critical radius* θ_{c} *of* M *is given by*

$$\tan^2 \theta_{\mathrm{c}} = \min\left\{ \inf_{x \neq \tilde{x},\, \psi(x) \in \mathrm{Int}\Gamma} \frac{(1-\alpha s)^2}{1 - s^2 - \alpha^2 t^2},\ \inf_{x \neq \tilde{x},\, \psi(x) \in \partial\Gamma} \frac{(1-\alpha s)^2}{1 - s^2 - \max\{0, \varepsilon(x)\alpha t\}^2} \right\},$$

where the infima are taken over $x, \tilde{x} \in \mathcal{X}$*, and* $\alpha \in [-1,1]$ *as well as additional conditions (arguments of* $\inf$*), and*

$$s = s(x,\tilde{x}) = \psi(x)^\top \psi(\tilde{x}), \quad t = t(x,\tilde{x}) = \frac{\psi_x(x)^\top \psi(\tilde{x})}{\|\psi_x(x)\|},$$

$\psi_x(x) = \partial\psi(x)/\partial x$,

$$\varepsilon(x) = \begin{cases} 1 & (\psi_x(x) \textit{ is inward to } \Gamma), \\ -1 & (\psi_x(x) \textit{ is outward to } \Gamma). \end{cases}$$

$\psi_x(x)$ is said to be inward or outward to Γ if the support cone of Γ at $\psi(x)$ is $S_{\psi(x)}\Gamma = \{\lambda\psi_x(s) \mid \lambda \geq 0\}$ or $\{\lambda\psi_x(s) \mid \lambda \leq 0\}$, respectively.

Theorem 4.2.5. *Assume Assumption 4.2.1. Moreover, assume that $\psi : \mathcal{X} \to \mathbb{S}^{p-1}$ is of C^4-class. Then, the local critical radius $\theta_{\mathrm{c,loc}}$ is given by*

$$\tan^2\theta_{\mathrm{c,loc}} = \min\left\{ \inf_{x\in\mathcal{X}:\kappa(x)\leq 2} \left\{1 - \frac{\kappa(x)}{4}\right\}, \inf_{x\in\mathcal{X}:\kappa(x)\geq 2} \frac{1}{\kappa(x)} \right\}$$

with

$$\kappa(x) = \frac{\psi_{xx}(x)^\top\psi_{xx}(x)}{\{\psi_x(x)^\top\psi_x(x)\}^2} - \frac{\{\psi_{xx}(x)^\top\psi_x(x)\}^2}{\{\psi_x(x)^\top\psi_x(x)\}^3} - 1, \tag{4.2.5}$$

where $\psi_x(x) = \partial\psi(x)/\partial x$ and $\psi_{xx}(x) = \partial^2\psi(x)/\partial x^2$.

The proofs of Theorems 4.2.4 and 4.2.5 are included in the Appendix.

4.3 A numerical example

In this section, we provide a numerical example to determine the accuracy of the approximation formula given in Theorem 4.2.2, and degree of conservativeness proved by Theorem 4.2.3.

Suppose that $f(x) = (1, x, x^2)^\top$, $\mathcal{X} = [-1, 1]$, and

$$\Sigma = \begin{pmatrix} 1 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & 1 \end{pmatrix}, \quad \Sigma^{1/2} = \begin{pmatrix} 1 & 0 & \frac{2}{3} \\ 0 & \sqrt{\frac{2}{3}} & 0 \\ 0 & 0 & \frac{\sqrt{5}}{3} \end{pmatrix}.$$

Then,

$$\psi(x) = \frac{1}{3(1+x^2)}\left(3 + 2x^2, \sqrt{6}x, \sqrt{5}x^2\right)^\top, \quad |\Gamma| = \int_{\mathcal{X}} \|\dot{\psi}(x)\|\,\mathrm{d}x = \int_{-1}^{1} \sqrt{\frac{2}{3}} \frac{1}{1+x^2}\,\mathrm{d}x = \frac{\pi}{\sqrt{6}}.$$

$\kappa(x)$ in (4.2.5) is always 5. Hence, the local critical radius is $\theta_{\mathrm{c,loc}} = \tan^{-1}(1/\sqrt{5}) = 0.134\pi$. Further, we can also confirm that the critical radius is the same as $\theta_{\mathrm{c}} = \theta_{\mathrm{c,loc}}$ using Mathematica (Wolfram Research, Inc., 2016).

Under this setting, we suppose the case of $k = 3$. The probability we need is

$$\Pr\left\{\max_{x\in[-1,1]} Y(x) \geq b^2\right\} = 1 - \Pr\bigl\{T(x,c) \leq b,\ \forall x \in [-1,1],\ \forall c \in \mathcal{C}\bigr\}, \tag{4.3.1}$$

where

$$Y(x) = \sum_{i=1}^{2}\{\xi_i^\top \psi(x)\}^2, \quad \xi_1, \xi_2 \sim \mathcal{N}_3(0, I) \text{ i.i.d.}$$

is a chi-square random process $Y(x)$ with two degrees of freedom. The tube formula for the upper tail probability (4.3.1) is

$$\overline{P}_{\text{tube}}(b) = \frac{\pi}{2\sqrt{6}}\bigl\{\overline{G}_3(b^2) - \overline{G}_1(b^2)\bigr\} + \overline{G}_2(b^2) = \left(\frac{\sqrt{\pi}}{2\sqrt{3}}b + 1\right)e^{-b^2/2}. \tag{4.3.2}$$

Figure 4.3.1 depicts the upper tail probability of the maximum (4.3.1) and its approximate value (4.3.2). We can see that the tube formula approximates the true upper tail probability with sufficient accuracy in the moderate tail regions (for example, the upper probability is less than 0.2), and it provides a conservative bound as per Theorem 4.2.3.

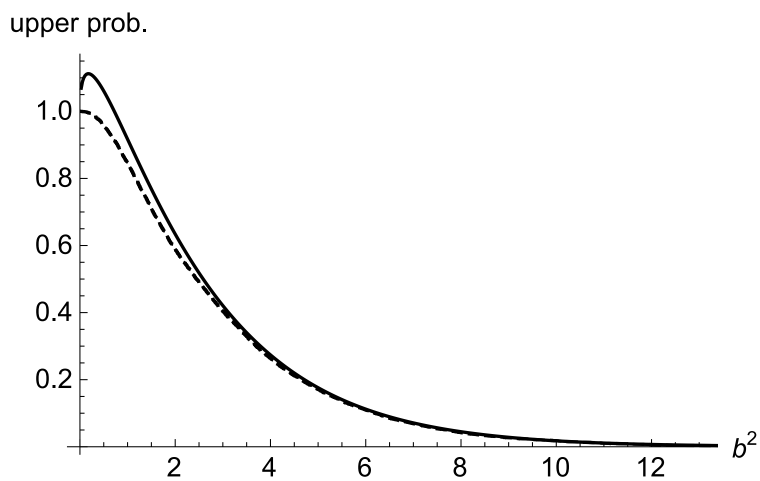

Figure 4.3.1: Upper tail probability of the maximum of chi-square process $Y(x)$.
(solid line: tube formula, dashed line: Monte Carlo with 10,000 replications)

We have proposed that the threshold for the confidence band should be determined as the solution $b = b_{\text{tube},1-\alpha}$ for $\overline{P}_{\text{tube}}(b) = \alpha$. Figure 4.3.2 depicts the actual confidence coefficient (coverage probability)

$$\Pr\left\{\max_{x\in[-1,1]} Y(x) \geq b^2_{\mathrm{tube},1-\alpha}\right\}, \quad \alpha \in [0,1].$$

This further demonstrates that the confidence bands obtained by the tube method are always conservative and very accurate.

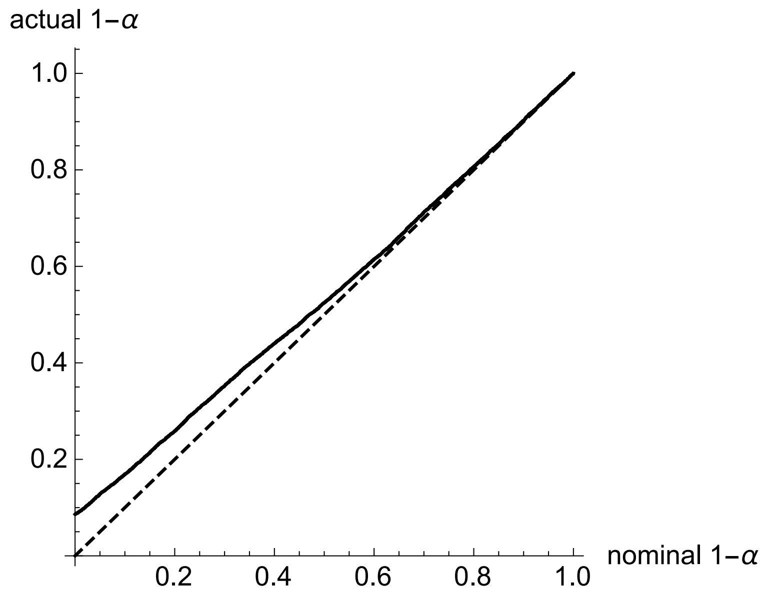

Figure 4.3.2: Nominal confidence coefficient vs. Actual confidence coefficient.
(solid line: actual confidence coefficient, dashed line: 45-degree line)

4.4 Simulation study under model misspecification

Throughout this section, it is assumed that the nonlinear model has a finite number of basis functions $g_i(x) = \beta_i^\top f(x)$ in (2.4.2). However, we can only approximate the true model in practice. Under a slight misspecification of the model, Sun and Loader (1994) estimated the bias of the coverage probability, and proposed an adjustment to the volume-of-tube formula. Although their approach may be applied to our model, the result would be more complicated. Instead, to investigate what happens under model misspecification, we conducted a Monte Carlo simulation study in the following setting.

The domain of explanatory variable is set to be $\mathcal{X} = [0,1]$. The data are generated from the model

$$y_{ij} = g_i(x_j) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0,1), \quad i = 1,\dots,k, \quad j = 1,\dots,n,$$

where $k = 3$, $n = 11$, and $x_j = (j-1)/n$, $j = 1,\dots,n$. As the true regression curve is $g_i(x)$, we assume three models.

Model 1:

$$g_i(x) = \beta_i^\top f_{2,5,0,1}(x), \quad \beta_1 = (0,\dots,0)^\top, \ \beta_2 = K(0,0,1/2,1,1)^\top, \ \beta_3 = K(0,0,4/3,0,0)^\top,$$

where $K = 1$, 3, or 9,

$$f_{d,m,a,b} = \left(B_d\left(\frac{x-a}{b-a}(m-d) - (i-d-1)\right)\right)_{i=1,\dots,m},$$

and $B_d(\cdot)$ is the B-spline function

$$B_d(x) = \sum_{r=0}^{d+1} (-1)^{d+1-r} \binom{d+1}{r} \frac{(r-x)_+^d}{d!} \tag{4.4.1}$$

(de Boor (1978), p. 89).

Model 2:

$$g_1(x) = 0, \quad g_2(x) = K\sin(x\pi/2), \quad g_3(x) = K\sin(x\pi), \quad K = 1,3,9.$$

Model 3:

$$g_1(x) = 0, \quad g_2(x) = K\frac{e^{-x/2} - e^{-x}}{e^{-1/2} - e^{-1}}, \quad g_3(x) = K\frac{\cosh(x-1/2) - 1}{\cosh(1/2) - 1}, \quad K = 1,3,9.$$

For all models, $g_2(x)$ is unimodal, and $g_3(x)$ is increasing. $g_2(x)$ and $g_3(x)$ are designed to have the range $[0, K]$.

We fit the curve $\beta_i^\top f_{2,m,0,1}(x)$ to the generated data y_{ij}, where $m = 3,\dots,10$. Using these models, we constructed a $1-\alpha = 0.95$ confidence band. Coverage probabilities were estimated based on Monte Carlo simulations with 1,000,000 replications, and are summarized in Table 4.4.1. In this table,

$$\delta = \max_{x\in\mathcal{X},\, c\in\mathcal{C}} \left| \frac{\sum_{i=1}^k c_i\{(\beta_i^*)^\top f_{2,m,0,1}(x) - g_i(x)\}}{\sqrt{f_{2,m,0,1}(x)^\top \Sigma f_{2,m,0,1}(x)}} \right|, \quad \Sigma = \left(\sum_{i=1}^n f_{2,m,0,1}(x_i)^\top f_{2,m,0,1}(x_i)\right)^{-1} \tag{4.4.2}$$

is the bias of regression function, where β_i^* is the best parameter in the assumed model $\beta_i^\top f_{2,m,0,1}(x)$.

$$\Delta = \max\{\alpha - \overline{P}_{\text{tube}}(b_{\text{tube},1-\alpha} + \delta), \overline{P}_{\text{tube}}(b_{\text{tube},1-\alpha} - \delta) - \alpha\} \tag{4.4.3}$$

is an approximate upper bound of the bias of coverage probability, where $b_{\text{tube},1-\alpha}$ is the approximate value of $b_{1-\alpha}$ obtained by the tube method. (See Appendix A.4 for the detail.)

From this table, we first see that, for the true models ($m = 5, 8$ when model 1 is true), the coverage probabilities are more than, but approximately equal to, the nominal value 0.95, meaning that the proposed method is valid. The most remarkable point is that, throughout the study, the coverage probabilities are kept at approximately 0.95, unless the assumed model is too small, and the bias δ is large.

Table 4.4.2 shows the average width of the confidence band defined by

$$W = \frac{\int_{\mathcal{X}} b_{\text{tube},1-\alpha} \sqrt{f(x)^\top \Sigma f(x)}\,\mathrm{d}x}{\int_{\mathcal{X}} \mathrm{d}x} = b_{\text{tube},0.95} \int_0^1 \sqrt{f_{2,m,0,1}(x)^\top \Sigma f_{2,m,0,1}(x)}\,\mathrm{d}x.$$

When the model is increasing in size, W is increasing in size. This suggests that a smaller model is preferable, unless it is too small to cause serious bias.

In summary, too small of a model should surely be avoided, whereas, a larger model has the disadvantage of having a wider confidence band. This trade-off is crucially important in practice, and a promising future research topic, although it is out of scope for this paper. For related topics, refer to Casella and Hwang (2012), for shrinkage confidence bands, and Leeb et al. (2015), for confidence band post-model selection.

Table 4.4.1: Coverage probability under model misspecification ($1-\alpha = 0.95$)
(prob: coverage probability, δ: bias (4.4.2), Δ: bound for coverage probability bias (4.4.3))

m	Model 1 ($K=1$)			Model 1 ($K=3$)			Model 1 ($K=9$)		
	prob	δ	Δ	prob	δ	Δ	prob	δ	Δ
3	0.9365	0.4692	0.1155	0.7872	1.4076	0.8240	0.0000	4.2227	1.3542
4	0.9422	0.3996	0.0965	0.8509	1.1987	0.6909	0.0076	3.5961	1.6907
5	0.9512	0.0000	0.0000	0.9512	0.0000	0.0000	0.9512	0.0000	0.0000
6	0.9511	0.1006	0.0176	0.9477	0.3018	0.0694	0.9111	0.9053	0.4550
7	0.9514	0.0448	0.0074	0.9509	0.1343	0.0251	0.9465	0.4030	0.1100
8	0.9515	0.0000	0.0000	0.9515	0.0000	0.0000	0.9515	0.0000	0.0000
9	0.9515	0.0175	0.0029	0.9514	0.0526	0.0090	0.9504	0.1578	0.0316
10	0.9516	0.0218	0.0036	0.9515	0.0653	0.0116	0.9508	0.1959	0.0421

m	Model 2 ($K=1$)			Model 2 ($K=3$)			Model 2 ($K=9$)		
	prob	δ	Δ	prob	δ	Δ	prob	δ	Δ
3	0.9509	0.06491	0.0099	0.9486	0.1947	0.0346	0.9277	0.5842	0.1640
4	0.9511	0.04999	0.0077	0.9498	0.1500	0.0264	0.9374	0.4499	0.1157
5	0.9512	0.01271	0.0019	0.9511	0.0381	0.0060	0.9504	0.1143	0.0199
6	0.9516	0.00494	0.0008	0.9516	0.0148	0.0023	0.9515	0.0445	0.0072
7	0.9514	0.00234	0.0004	0.9514	0.0070	0.0011	0.9514	0.0211	0.0034
8	0.9515	0.00137	0.0002	0.9515	0.0041	0.0007	0.9515	0.0123	0.0020
9	0.9515	0.00119	0.0002	0.9515	0.0036	0.0006	0.9515	0.0107	0.0017
10	0.9516	0.00076	0.0001	0.9516	0.0023	0.0004	0.9516	0.0068	0.0011

m	Model 3 ($K=1$)			Model 3 ($K=3$)			Model 3 ($K=9$)		
	prob	δ	Δ	prob	δ	Δ	prob	δ	Δ
3	0.9512	0.02384	0.0034	0.9508	0.07152	0.0109	0.9483	0.2146	0.0390
4	0.9513	0.00766	0.0011	0.9512	0.02299	0.0034	0.9511	0.0690	0.0109
5	0.9512	0.00218	0.0003	0.9512	0.00653	0.0010	0.9512	0.0196	0.0030
6	0.9516	0.00095	0.0002	0.9516	0.00285	0.0004	0.9516	0.0086	0.0013
7	0.9514	0.00046	0.0001	0.9514	0.00138	0.0002	0.9514	0.0042	0.0006
8	0.9515	0.00028	0.0000	0.9514	0.00083	0.0001	0.9515	0.0025	0.0004
9	0.9515	0.00024	0.0000	0.9515	0.00071	0.0001	0.9515	0.0021	0.0003
10	0.9516	0.00014	0.0000	0.9516	0.00042	0.0001	0.9516	0.0013	0.0002

Table 4.4.2: Average band-width W $(1 - \alpha = 0.95)$

m	3	4	5	6	7	8	9	10
W	1.463	1.752	2.017	2.275	2.546	2.764	2.990	3.211

Chapter 5

Analysis of Growth Curve

5.1 Growth curve

In this chapter we demonstrate the analysis of mouse growth as an illustration. Sun, Raz, and Faraway (1999) proposed simultaneous confidence bands for a growth curve by virtue of the volume-of-tube method. Differently from their analysis, we focus on the contrast of several growth curves.

Mice are one of the most popular model organisms, and are often used in genomic research. Figure 5.1.1 depicts the average body weights of male mice from four different strains measured from 2 to 20 weeks after birth. The four strains are C57BL/6 (referred to as B6), MSM/Ms (MSM), B6-Chr17$^{\mathrm{MSM}}$(B6-17), and B6-ChrXT$^{\mathrm{MSM}}$(B6-XT). Among these, B6 is the most common laboratory strain and serves as the standard. MSM is a wild-derived strain having contrasting properties to B6 such as non-black color, small size, and aggressive behavior. B6-17 and B6-XT are artificial strains known as consomic mice made from B6 and MSM. B6-17 has all the chromosomes from B6, and only chromosome 17 from MSM; B6-XT has all the chromosomes from B6, and only half of the X chromosome from MSM. By comparing the consomic strains with B6, we expect to reveal the role of each chromosome.

The dataset we utilized is publicly available as Supplemental Table S1 of Takada et al. (2008). In their experiments, the weight (unit: gram) y_{ijh} of the hth individual from strain i was measured at time point x_j. The measurement time points were $\{x_1, \ldots, x_{10}\} = \{2, 4, \ldots, 20\}$ $(n = 10)$. This dataset includes the average body weight y_{ij} of strain i at time x_j, and its standard error

$$y_{ij} = \frac{1}{r_i} \sum_{h=1}^{r_i} y_{ijh}, \quad \widehat{\mathrm{s.e.}}(y_{ij}) = \sqrt{\frac{1}{r_i^2} \sum_{h=1}^{r_i} (y_{ijh} - y_{ij})^2},$$

as well as the number r_i of individuals of strain i.

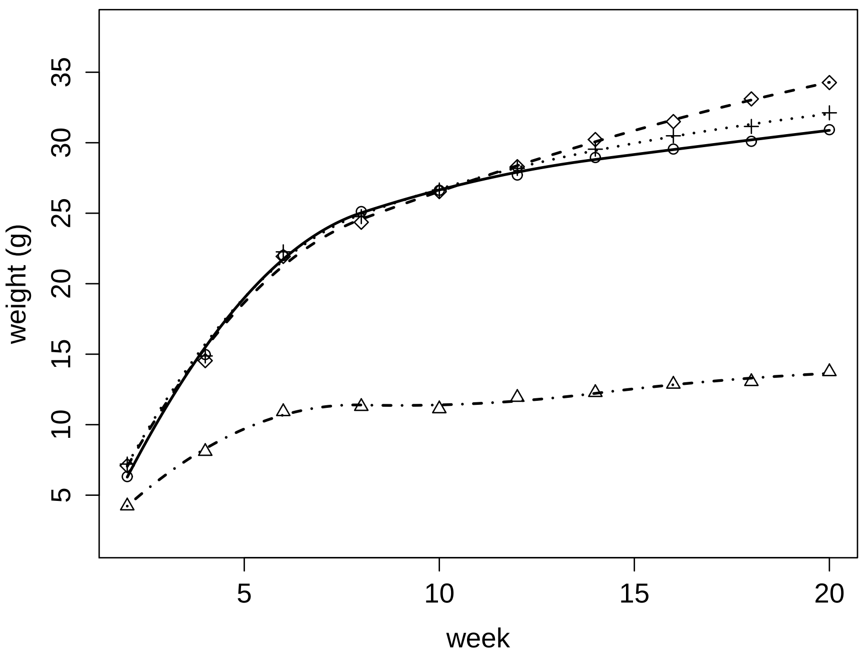

Figure 5.1.1: Average body weights of mice from four strains.
(sample mean: ○ (B6), + (B6-17), ◇ (B6-XT), △ (MSM);
fitted curve: — (B6), ··· (B6-17), −− (B6-XT), −·− (MSM))

In the following analysis, we use $k = 3$ groups (strains) B6 ($i = 1$), B6-17 ($i = 2$), and B6-XT ($i = 3$). The number of individuals are $r_1 = 12$, $r_2 = 24$, and $r_3 = 12$.

We fit the model (2.4.1) to these data. We estimate the variance as

$$\widehat{\sigma}(x_j)^2 = \frac{1}{\sum_{i=1}^{k}(r_i - 1)} \sum_{i=1}^{k} \sum_{h=1}^{r_i} (y_{ijh} - y_{ij})^2 = \frac{1}{\sum_{i=1}^{k}(r_i - 1)} \sum_{i=1}^{k} r_i^2 \widehat{\text{s.e.}}(y_{ij})^2,$$

which is used as the true value $\sigma(x_j)^2$ hereafter. Figure 5.1.2 plots the estimated standard error $\widehat{\sigma}(x_j)$. One particular feature of this dataset is that the experiment is well controlled and measurement errors are quite small.

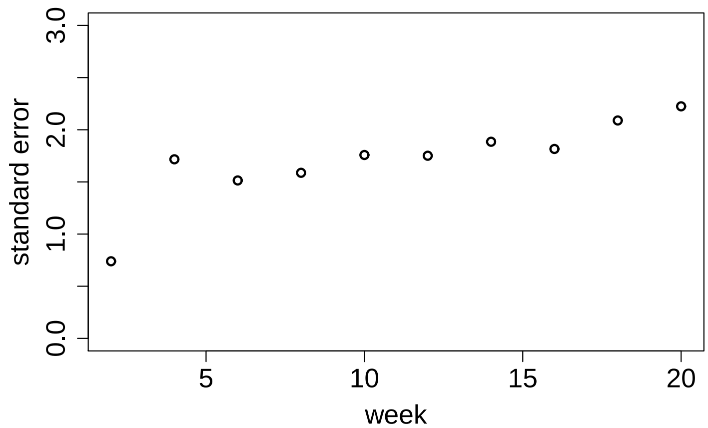

Figure 5.1.2: Estimated standard error $\widehat{\sigma}(x_j)$.

5.2 Model selection

As the basis function $f(x)$, we consider a family of basis functions

$$f(x) = f_{d,m,2,20}(x) = \left(B_d\left(\frac{x-2}{20-2}(m-d) - (i-d-1) \right) \right)_{1 \le i \le m},$$

with $B_d(x)$ given in (4.4.1). $f_{d,m,2,20}(x)$ consists of m B-spline bases with equally-spaced knots at intervals of $(20-2)/(m-d)$. Note that $f_{d,m,2,20}(x)$ is piecewise of class C^d.

In the range $d = 2, 3, 4$ and $m = d+1, d+2, \ldots, n\,(= 10)$, we searched for the best model that minimizes AIC and BIC defined below:

$$\mathrm{AIC}_{d,m} = L_{d,m} + 2km, \quad \mathrm{BIC}_{d,m} = L_{d,m} + \sum_{i=1}^{k} \ln(r_i n) m, \quad L_{d,m} = \sum_{i=1}^{k} r_i \sum_{j=1}^{n} \frac{(y_{ij} - \widehat{y}_{ij})^2}{\sigma(x_j)^2}$$

with $k = 3$, $n = 10$, where $\widehat{y}_{ij} = \widehat{\beta}_i^\top f_{d,m,2,20}(x_j)$. In both criteria, the minimizer was $(d, m) = (2, 5)$, which we use as the true value hereafter.

Suppose that we are interested in the period $\mathcal{X} = [a, b] = [2, 20]$. An approximate value of the length of Γ in (2.5.1) is given by

$$|\Gamma| \approx \sum_{t=1}^{N} \left\| \psi(x_t) - \psi(x_{t-1}) \right\|,$$

where $x_t = a + t(b-a)/N$, $t = 0, 1, \ldots, N$. When $N = 10,000$, the approximate value of $|\Gamma|$ is $6.989 = 2.225\pi$. Using this, the critical value is $b_{1-\alpha} = 3.258$ ($\alpha = 0.05$).

5.3 Difference in body weight

To compare k groups, various types of contrasts are used. For a pairwise comparison between group i and group j, we choose $c = (\ldots, 0, \underset{i\text{th}}{1}, 0, \ldots, 0, \underset{j\text{th}}{-1}, 0, \ldots)$. For the comparison of groups $\{i, j\}$ and group k, we use

$$c = \left(\ldots, 0, \underset{i\text{th}}{\frac{r_i}{r_i + r_j}}, 0, \ldots, 0, \underset{j\text{th}}{\frac{r_j}{r_i + r_j}}, 0, \ldots, 0, \underset{k\text{th}}{-1}, 0, \ldots \right).$$

Figure 5.3.1 depicts the difference curves of strains B6-17 vs. B6 (left) and B6-XT vs. B6 (right), and their 95% simultaneous confidence bands. In the left panel, the horizontal line representing zero difference is almost between the confidence bands. This indicates that there is no significant difference between B6-17 and B6. In contrast, in the right panel, after around week 14, the horizontal line is outside the confidence bands, thereby indicating that B6-XT and B6 are different during this period.

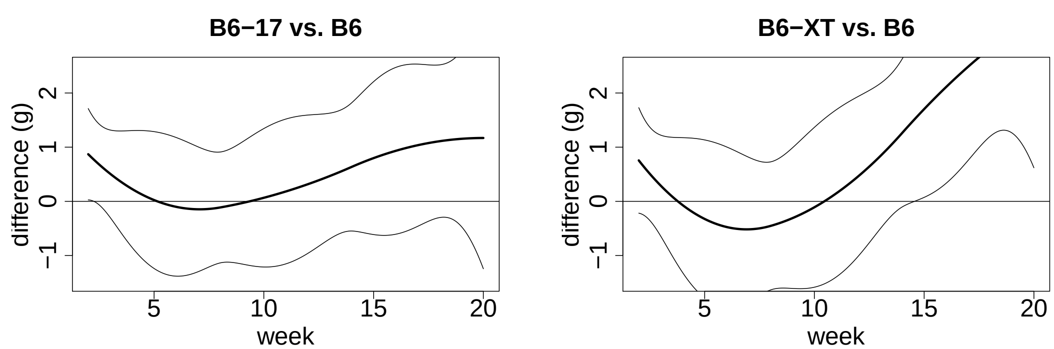

Figure 5.3.1: Differences of body weights and 95% confidence bands.

For a fixed x, the test statistic for the null hypothesis $H_{0,x} : \beta_1^\top f(x) = \ldots =$

$\beta_k^\top f(x)$ is

$$\chi^2(x) = \frac{1}{f(x)^\top \Sigma f(x)} \sum_{i=1}^{k} r_i \left\{ \widehat{\beta}_i^\top f(x) - \frac{\sum_{i=1}^{k} r_i \widehat{\beta}_i^\top f(x)}{\sum_{i=1}^{k} r_i} \right\}^2 .$$

For a fixed x, the null distribution is the chi-square distribution with $k-1$ degrees of freedom. However, for the overall null hypothesis $H_0 : \beta_1^\top f(x) = \ldots = \beta_k^\top f(x)$ for all $x \in \mathcal{X}$, the distribution of the maximum of the chi-square random process should be used. Figure 5.3.2 shows $\chi^2(x)$ and its upper 5% critical value $b_{0.95}^2$. As already shown in Figure 5.3.1, after around week 14, the hypothesis of equality is rejected.

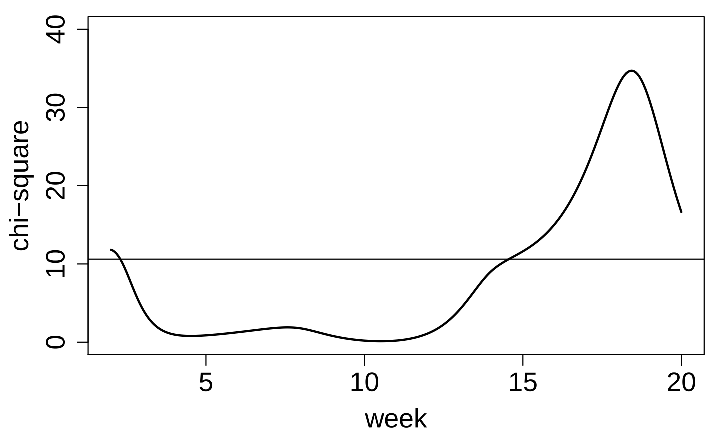

Figure 5.3.2: Chi-square process $\chi^2(x)$ and its upper 5% critical value.

Throughout this paper, it is assumed that the nonlinear model (2.4.2) is known to be true in advance. However, we hardly know the true model in practice. Under a slight misspcification of the model, Sun and Loader (1994) estimated the bias of the confidence probability, and proposed an adjustment to the volume-of-tube formula. Although their approach may be applied to our model, the result could be more complicated. This would have to form part of our future research.

5.4 Simulation study under model misspecifications

To investigate what happens under model misspecifications, we conducted a Monte Carlo simulation study in the following setting: In section 5, we obtained estimates

$\widehat{\beta}_i$, $i = 1, 2, 3$, under the model with $f(x) = f_{2,5}(x)$. Here, by assuming the estimates $\widehat{\beta}_i^\top f_{2,5}(x_j)$ and $\widehat{\sigma}(x_j)^2$ to be the true values for $\mathrm{E}[y_{ijh}]$ and $\mathrm{Var}(y_{ijh})$, we generate the data y_{ijh}, $i = 1, \ldots, k\,(= 3)$, $j = 1, \ldots, n\,(= 10)$, and $h = 1, \ldots, r_i$, and then fit the models with $f(x) = f_{2,m}(x)$, $m = 3, 4, \ldots, 10$.

The model with $m = 5$ is a true model. The model with $m = 8$ is also a true model, since the set of knots for $m = 8$ is $\{2, 5, 8, 11, 14, 17, 20\}$, which includes the set of knots $\{2, 8, 14, 20\}$ for $m = 5$. On the other hand, the models with $m = 3, 4, 6, 7, 9, 10$ are not true and should have some biases.

Figure 5.4.1 depicts the simultaneous confidence probabilities estimated by Monte Carlo simulation with 1,000,000 replications per point. We first see that for the true models ($m = 5, 8$), the confidence probabilities are more than but almost equal to the nominal values, which means that the proposed method is valid. Including these cases, when $m \geq 5$, the confidence probabilities are almost equal to the nominal values. For the case $m = 4$, a slight bias is observed, whereas for the case $m = 3$, serious bias occurs.

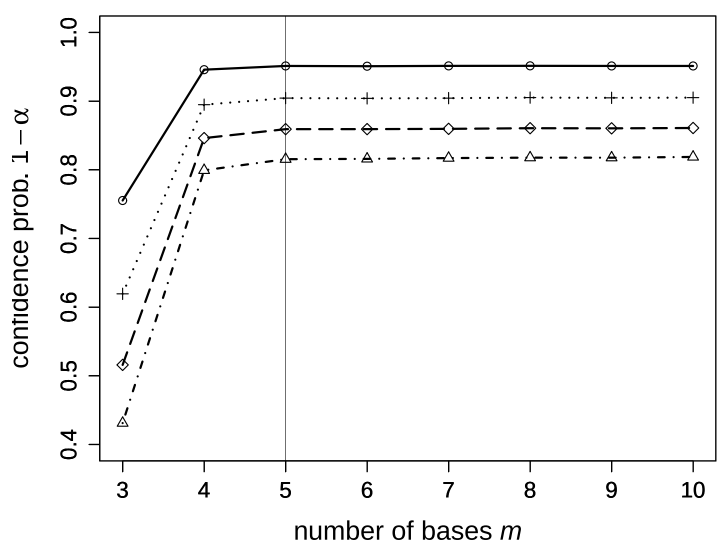

Figure 5.4.1: Confidence probability $1 - \alpha$ under basis vector $f_{2,m}$. True model: $m = 5$.

(Nominal confidence coefficient 0.95: —○—, 0.90: $\cdots + \cdots$, 0.85: −−◇−−, 0.80: −·−△−·−)

On the other hand, since m is the dimension of parameters to be estimated, the smaller m is, the narrower the bandwidth is. Figure 5.4.2 depicts the average bandwidth of the $1-\alpha$ simultaneous confidence bands:

$$|\mathcal{X}|^{-1}\int_{\mathcal{X}} b_\alpha \sqrt{f(x)^\top \Sigma f(x)}\mathrm{d}x, \quad |\mathcal{X}| = \int_{\mathcal{X}} \mathrm{d}x.$$

Considering the observation that the model with $m = 4$ yields shorter bandwidth and has almost no bias, the model with $m = 4$ may be more favorable than the true model with $m = 5$ in practice.

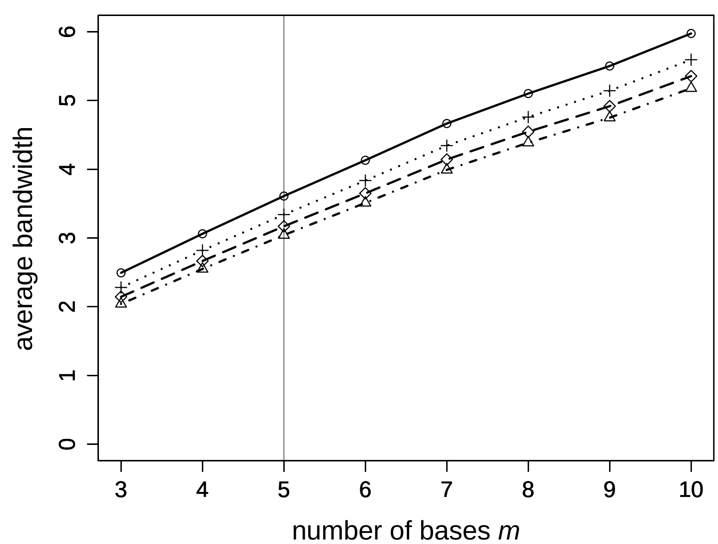

Figure 5.4.2: Average bandwidth under basis vector $f_{2,m}$. True model: $m = 5$.

(Nominal confidence coefficient 0.95 : —○—, 0.90 : ···+···, 0.85 : −−◇−−, 0.80 : −·−△−·−)

In summary, extensive model misspecification (i.e., $m = 3$) causes serious bias, and should surely be avoided. Whereas a smaller and slightly misspecified model (i.e., $m = 4$) may lead to good performance, i.e., almost no bias and narrower bandwidth. The use of a lower dimensional model reminds us of a confidence band based on a shrinkage estimator, which is still under development (Casella and Hwang, 2012). The construction of optimized simultaneous confidence bands using a lower dimensional and possibly misspecified model might be a challenging future research topic

Chapter 6

Conclusion

In this study, we review simultaneous confidence bands and the volume-of-tube method. In the literature, there are several works on confidence bands for linear regression models, such as Working and Hotelling (1929), Hotelling (1951), Spurrier (1999), and Jamshidian, Liu, and Bretz (2010). Unlike these studies, we consider the construction of simultaneous bands for all of the contrasts between several nonlinear regression models. The tube formula is given in a mathematical form via the volume-of-tube method. The critical value of a confidence band is determined from the distribution of the maximum of a chi-square random process defined on the domain of the explanatory variables. We use the volume-of-tube method to derive an upper tail probability formula for the maximum of a chi-square random process, which is sufficiently accurate in commonly used tail regions. Moreover, we prove that the formula obtained is equivalent to the expectation of the Euler-Poincaré characteristic of the excursion set of the chi-square random process and, hence, is conservative. This result is therefore a generalization of Naiman's inequality for Gaussian random processes. We provide a numerical example to address this method. The simulation study under model misspecification is given thereafter. Growth curves of consomic mice are analyzed as an illustrative example.

As a future research topic, we consider another application of the volume-of tube method on the statistical parametric maps (SPMs) problem. Statistical parametric mapping is an important tool to detect significantly activated regions of cerebral tissue. The threshold to render the probability of one or more activated regions of one voxel or larger is suitably small (e.g., 0.05). Friston et al. (1994) present an approximate analysis giving the probability that one or more activated regions of or larger than a specified volume could have occurred by chance.

SPMs are spatially extended statistical processes that are used to test hypotheses about regionally specific effects in neuroimaging data. The voxel values of SPMs are distributed according to some known probability density function under the null hypothesis (Friston et al., 1990). Any continuous probability density function can be transformed to the Gaussian distribution or a z-statistic. If the degrees of freedom of the original distribution are reasonably high, the resulting SPM approximates a Gaussian field or an SPMz.

We define the following three characteristics of an SPM:

(1) The number (N) of voxels above a threshold (the number of voxels in the excursion set that have values greater than a threshold)

(2) The number (m) of activated regions (clusters or connected subsets of the excursion set)

(3) The number (n) of voxels in each of these clusters.

Each of these numbers has its own probability density function: $\Pr(N = x)$, $\Pr(m = x)$, and $\Pr(n = x)$, respectively. These probability functions provide a fairly complete characterization of the SPM and allow a number of hypotheses to be addressed. The particular probability we are interested in is the probability of obtaining at least one activation with k or more voxels. This is the same as the probability that the largest region has k or more voxels = $\Pr(n_{\max} \geq k)$, where $n_{\max}$ is the number of voxels in the largest region. Friston et al. (1992) solves this problem by using the Euler characteristic, whereas we will use the volume-of-tube method. Here are two points that need to be reconsidered. First, the assumptions that the image data has stationarity and that it follows a Gaussian random field are not realistic since brain image data is completely different from a Gaussian random field. Second, the formula given by the Euler characteristic is highly simplified, which may cause low accuracy. We try to give a more complex formula with higher accuracy by using the volume-of-tube method.

Appendix A

Appendix: Proofs

A.1 Proof of Theorem 4.2.2

Contribution of the inner points $\mathrm{Int}\,M$

Here, we obtain the coefficients w_{d+1-e} in (3.3.3) when M is given in (4.2.1).

Let $h = h(\theta)$, $\theta = (\theta_i)_{1\le k-2}$, be a local coordinate system of $\mathbb{S}^{k-2}$. For example,

$$h = h(\theta) = \begin{pmatrix} \cos\theta_1 \\ \sin\theta_1\cos\theta_2 \\ \sin\theta_1\sin\theta_2\cos\theta_3 \\ \vdots \\ \sin\theta_1\ldots\sin\theta_{k-3}\cos\theta_{k-2} \\ \sin\theta_1\cdots\sin\theta_{k-3}\sin\theta_{k-2} \end{pmatrix}_{(k-1)\times 1},$$

where

$$\theta \in \Theta = \{(\theta_1,\ldots,\theta_{k-2}) \mid 0 \le \theta_i \le \pi \ (i=1,\ldots,k-3),\ 0 \le \theta_{k-2} < 2\pi\}.$$

Let $(x,\theta) \in \mathcal{X}\times\Theta$ be fixed, and let $\phi(x,\theta) = h(\theta)\otimes\psi(x) \in M$. We write $\psi = \psi(x)$, $h = h(\theta)$ and $\phi = \phi(x,\theta)$ for simplicity. We first assume that $x \in \mathrm{Int}\mathcal{X}$, hence, $\phi(x,\theta) \in \mathrm{Int}M$.

By applying the Gram-Schmidt orthonormalization to the sequence $\psi, \partial\psi/\partial x, \partial^2\psi/\partial x^2, \ldots$, we construct the orthonormal basis (ONB) $\psi_{(i)}$, $i = 0,\ldots,p-1$, of $\mathbb{R}^p$. The first three bases are

$$\psi_{(0)} = \psi, \quad \psi_{(1)} = \frac{1}{\sqrt{g}}\frac{\partial\psi}{\partial x}, \quad \psi_{(2)} = \frac{1}{\sqrt{\eta - \frac{\gamma^2}{g} - g^2}}\left(\frac{\partial^2\psi}{\partial x^2} + g\psi - \frac{\gamma}{g}\frac{\partial\psi}{\partial x}\right),$$

where

$$g = g(x) = \left(\frac{\partial\psi}{\partial x}\right)^{\top}\left(\frac{\partial\psi}{\partial x}\right), \quad \gamma = \gamma(x) = \left(\frac{\partial^2\psi}{\partial x^2}\right)^{\top}\left(\frac{\partial\psi}{\partial x}\right), \quad \eta = \eta(x) = \left(\frac{\partial^2\psi}{\partial x^2}\right)^{\top}\left(\frac{\partial^2\psi}{\partial x^2}\right).$$

Similarly, from the sequence h, $\partial h/\partial\theta_i$, $i = 1,\ldots,k-2$, we obtain ONB $h_{(i)}$, $i = 0,\ldots,k-2$, of $\mathbb{R}^{k-1}$. We prepare a $(k-2)\times(k-2)$ upper triangle matrix D such that

$$\left(h, \frac{\partial h}{\partial\theta_1}, \ldots, \frac{\partial h}{\partial\theta_{k-2}}\right) = \left(h_{(0)}, h_{(1)}, \ldots, h_{(k-2)}\right)\begin{pmatrix} 1 & 0 \\ 0 & D \end{pmatrix}, \quad \text{or } D = \left(h_{(i)}^{\top}\frac{\partial h}{\partial\theta_j}\right)_{1\le i,j\le k-2}.$$

Now we have the ONB $h_{(i)}\otimes\psi_{(j)}$, $i = 0,\ldots,k-2$, $j = 0,\ldots,p-1$, of the ambient space $\mathbb{R}^n$ with $n = p(k-1)$. Note that $\phi = h_{(0)}\otimes\psi_{(0)}$.

The tangent space $T_\phi M$ is spanned by

$$\frac{\partial\phi}{\partial x} = h\otimes\frac{\partial\psi}{\partial x}, \quad \frac{\partial\phi}{\partial\theta_i} = \frac{\partial h}{\partial\theta_i}\otimes\psi, \quad i = 1,\ldots,k-2.$$

The metric matrix of $T_\phi M$ with respect to the parameter $x, \theta_1, \ldots, \theta_{k-2}$ is

$$\begin{pmatrix} g & 0 \\ 0 & G \end{pmatrix}, \quad \text{where } G = \left(\left(\frac{\partial h}{\partial\theta_i}\right)^{\top}\left(\frac{\partial h}{\partial\theta_j}\right)\right)_{1\le i,j\le k-2} = D^{\top}D. \tag{A.1.1}$$

$T_\phi M$ has the ONB $h_{(0)}\otimes\psi_{(1)}$, $h_{(i)}\otimes\psi_{(0)}$, $i = 1,\ldots,k-2$.

The normal space perpendicular to $T_\phi(\mathrm{co}(M)) = T_\phi M\oplus\mathrm{span}\{\phi\}$ is

$$\begin{aligned} N_\phi(\mathrm{co}(M)) = \mathrm{span}\big\{& h_{(0)}\otimes\psi_{(j)},\ j = 2,\ldots,p-1; \\ & h_{(i)}\otimes\psi_{(j)},\ i = 1,\ldots,k-2,\ j = 1,\ldots,p-1\big\}. \end{aligned} \tag{A.1.2}$$

The second order derivatives of $\phi = \phi(x,\theta)$ are

$$\frac{\partial^2\phi}{\partial x^2} = h\otimes\frac{\partial^2\psi}{\partial x^2}, \quad \frac{\partial^2\phi}{\partial x\partial\theta_i} = \frac{\partial h}{\partial\theta_i}\otimes\frac{\partial\psi}{\partial x}, \quad \frac{\partial^2\phi}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 h}{\partial\theta_i\partial\theta_j}\otimes\psi.$$

Taking the inner product of the second derivatives and the ONB of $N_\phi(\mathrm{co}(M))$ listed in (A.1.2), we see that the nonzero elements of the second fundamental form are

$$-\left(h\otimes\frac{\partial^2\psi}{\partial x^2}\right)^{\top}\left(h_{(0)}\otimes\psi_{(2)}\right) = -\left(\frac{\partial^2\psi}{\partial x^2}\right)^{\top}\psi_{(2)} = -\zeta,$$

where

$$\zeta = \zeta(x) = \sqrt{\eta - \frac{\gamma^2}{g} - g^2}$$

and

$$-\left(\frac{\partial h}{\partial \theta_i} \otimes \frac{\partial \psi}{\partial x}\right)^\top (h_{(j)} \otimes \psi_{(1)}) = -\left(\frac{\partial h}{\partial \theta_i}\right)^\top h_{(j)}\sqrt{g} = -D_{ji}\sqrt{g}.$$

We renumber the ONB of $N_\phi(\mathrm{co}(M))$ as

$$N_1 = h_{(0)} \otimes \psi_{(2)}, \quad N_i = h_{(i-1)} \otimes \psi_{(1)},\ i = 2, \ldots, k-1,$$

and $N_k, \ldots, N_{pk-p-k}$ are the other vectors. Write $N(t) = \sum_{i=1}^{pk-p-k} N_i t_i$, where $t = (t_1, \ldots, t_{pk-p-k})$. Then,

$$\begin{aligned}
-\left(\frac{\partial^2 \phi}{\partial x^2}\right)^\top N(t) &= -\,\zeta t_1, \\
-\left(\frac{\partial^2 \phi}{\partial x \partial \theta_i}\right)^\top N(t) &= -\sum_{j=1}^{k-2} D_{ji} t_{j+1}\sqrt{g}, \\
-\left(\frac{\partial^2 \phi}{\partial \theta_i \partial \theta_j}\right)^\top N(t) &= 0.
\end{aligned}$$

Therefore, the second fundamental form (unnormalized version) in the direction $N(t)$ is

$$\begin{pmatrix} -\zeta t_1 & -(t_2, \ldots, t_{k-1}) D\sqrt{g} \\ -D^\top \begin{pmatrix} t_2 \\ \vdots \\ t_{k-1} \end{pmatrix} \sqrt{g} & 0 \end{pmatrix}. \tag{A.1.3}$$

Multiplication of the inverse of the metric (A.1.1) enables us to obtain the normalized version of the second fundamental form. Noting that

$$\begin{pmatrix} g & 0 \\ 0 & G \end{pmatrix}^{-1} = \begin{pmatrix} g & 0 \\ 0 & D^\top D \end{pmatrix}^{-1} = \begin{pmatrix} 1/\sqrt{g} & 0 \\ 0 & D^{-1} \end{pmatrix} \begin{pmatrix} 1/\sqrt{g} & 0 \\ 0 & (D^\top)^{-1} \end{pmatrix},$$

we multiply

$$\begin{pmatrix} 1/\sqrt{g} & 0 \\ 0 & (D^\top)^{-1} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1/\sqrt{g} & 0 \\ 0 & D^{-1} \end{pmatrix}$$

from the left and right to (A.1.3), respectively, to obtain

$$\begin{pmatrix} 1/\sqrt{g} & 0 \\ 0 & (D^\top)^{-1} \end{pmatrix} \begin{pmatrix} -\zeta t_1 & -(t_2,\dots,t_{k-1})D\sqrt{g} \\ -D^\top \begin{pmatrix} t_2 \\ \vdots \\ t_{k-1} \end{pmatrix} \sqrt{g} & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{g} & 0 \\ 0 & D^{-1} \end{pmatrix}$$
$$= \begin{pmatrix} -(\zeta/g)t_1 & -(t_2,\dots,t_{k-1}) \\ -\begin{pmatrix} t_2 \\ \vdots \\ t_{k-1} \end{pmatrix} & 0 \end{pmatrix} = H(x,\theta;N(t)).$$

This is the second fundamental form with respect to the orthonormal coordinates. Now we have

$$\mathrm{tr}_e H(x,\theta;N(t)) = \begin{cases} 1 & (e=0), \\ -(\zeta/g)t_1 & (e=1), \\ -\sum_{j=2}^{k-1} t_j^2 & (e=2), \\ 0 & (\text{otherwise}). \end{cases} \tag{A.1.4}$$

Next, we evaluate the integral

$$\int_{v\in N_\phi(\mathrm{co}(M))\cap\mathbb{S}^{n-1}} \mathrm{tr}_e H(x,\theta;v)\,\mathrm{d}v, \tag{A.1.5}$$

where $n = p(k-1)$, $\mathrm{d}v$ is the volume element of $N_\phi(\mathrm{co}(M))\cap\mathbb{S}^{n-1}$, by following Section 4.2.2 of Kuriki and Takemura (2001). Recall that $d = \dim M = k-1$.

Because $N_\phi(\mathrm{co}(M))$ is a linear space of dimension $n-d-1 = p(k-1)-(k-1)-1 = pk-p-k$, $N_\phi(\mathrm{co}(M))\cap\mathbb{S}^{n-1}$ is nothing but a $(pk-p-k-1)$-dimensional unit sphere. Hence,

$$\int_{v\in N_\phi(\mathrm{co}(M))\cap\mathbb{S}^{n-1}} \mathrm{d}v = \mathrm{Vol}(\mathbb{S}^{pk-p-k-1}) = \Omega_{pk-p-k}.$$

Therefore, if V is distributed as the uniform distribution on $N_\phi M\cap\mathbb{S}^{n-1}$, denoted by $\mathrm{Unif}(N_\phi M\cap\mathbb{S}^{n-1})$, then (A.1.5) $= \Omega_{pk-p-k}\times \mathrm{E}\{\mathrm{tr}_e H(x,\theta;V)\}$.

Suppose that $T = (T_1,\dots,T_{pk-p-k}) \sim \mathcal{N}_{pk-p-k}(0,I)$, and let $N(T) = \sum_{i=1}^{pk-p-k} N_i T_i$. Then,

$$\|N(T)\|^2 = \sum_{i=1}^{pk-p-k} T_i^2 \sim \chi^2_{pk-p-k} \quad \text{and} \quad V = \frac{N(T)}{\|N(T)\|} \sim \mathrm{Unif}(N_\phi M\cap\mathbb{S}^{n-1})$$

are independently distributed. Hence,

$$\mathrm{E}\{\mathrm{tr}_e H(x,\theta;N(T))\} = \mathrm{E}\{\|N(T)\|^e \mathrm{tr}_e H(x,\theta,V)\} = \mathrm{E}\{\|N(T)\|^e\}\mathrm{E}\{\mathrm{tr}_e H(x,\theta,V)\},$$

and

$$\mathrm{E}\{\mathrm{tr}_e H(x,\theta,V)\} = \frac{\mathrm{E}\{\mathrm{tr}_e H(x,\theta;N(T))\}}{\mathrm{E}\big\{(\chi^2_{pk-p-k})^{e/2}\big\}}.$$

From (A.1.4),

$$\mathrm{E}\{\mathrm{tr}_e H(x,\theta;N(T))\} = \begin{cases} 1 & (e=0), \\ 0 & (e=1), \\ -\sum_{i=2}^{k-1} \mathrm{E}(T_i^2) = -(k-2) & (e=2), \\ 0 & (\text{otherwise}), \end{cases}$$

hence,

$$\mathrm{E}\{\mathrm{tr}_e H(x,\theta;V)\} = \begin{cases} 1 & (e=0), \\ -\dfrac{k-2}{pk-p-k} & (e=2), \\ 0 & (\text{otherwise}). \end{cases}$$

Therefore,

$$\int_{v\in N_\phi(\mathrm{co}(M))\cap\mathbb{S}^{n-1}} \mathrm{tr}_e H(x,\theta,v)\,\mathrm{d}v = \begin{cases} \Omega_{pk-p-k} & (e=0), \\ -\dfrac{k-2}{pk-p-k}\Omega_{pk-p-k} & (e=2), \\ 0 & (\text{otherwise}). \end{cases}$$

Note that the results are independent of x and θ. This implies that the integral in (3.3.3) with respect to the volume element

$$\mathrm{d}u = \sqrt{g}\,\mathrm{d}x \times |G(\theta)|^{1/2}\,\mathrm{d}\theta_1 \ldots \mathrm{d}\theta_{k-2}$$

is simply multiplying the constant

$$\mathrm{Vol}(M) = |\Gamma| \times \mathrm{Vol}(\mathbb{S}^{k-2}) = |\Gamma|\Omega_{k-1}.$$

Finally, from (3.3.3) of Proposition 3.3.1,

$$w_{d+1} = w_k = \frac{1}{\Omega_k \Omega_{pk-p-k}} \times |\Gamma| \Omega_{k-1} \Omega_{pk-p-k},$$
$$w_{d-1} = w_{k-2} = \frac{1}{\Omega_{k-2} \Omega_{pk-p-k+2}} \times |\Gamma| \Omega_{k-1} \times \left(-\frac{k-2}{pk-p-k} \right) \Omega_{pk-p-k},$$

and the other w's are zero. Simple calculations give

$$w_k = -w_{k-2} = \frac{\Gamma(\frac{k}{2})}{\sqrt{\pi}\Gamma(\frac{k-1}{2})} |\Gamma|. \tag{A.1.6}$$

Contribution of the boundary ∂M

We obtain here the coefficients w'_{d-e} in (3.3.4) when M is given in (4.2.1).

Suppose that $\mathcal{X} = [a, b]$. Then,

$$\partial M = \{\phi(a, \theta) \mid \theta \in \Theta\} \sqcup \{\phi(b, \theta) \mid \theta \in \Theta\}.$$

Let $x = a$ and $\theta \in \Theta$ be fixed. Then, $\phi(a, \theta) \in \partial M$. The metric of the boundary ∂M at $\phi(x, \theta)$ is

$$\left(\frac{\partial \phi}{\partial \theta_i} \right)^\top \left(\frac{\partial \phi}{\partial \theta_j} \right) \bigg|_{(a,\theta)} = (G(\theta))_{ij}.$$

Note that $\partial(\mathrm{co}(M)) = \mathrm{co}(\partial M)$. The support cone of $\mathrm{co}(M)$ at $\phi(a, \theta) \in \partial(\mathrm{co}(M))$ is

$$S_{\phi(a,\theta)}(\mathrm{co}(M)) = L \oplus K_1,$$

where

$$L = \mathrm{span} \left\{ h \otimes \psi,\ \frac{\partial h}{\partial \theta_i} \otimes \psi,\ i = 1, \dots, k-2 \right\}, \quad K_1 = \left\{ \lambda \left(h \otimes \frac{\partial \psi}{\partial x} \right) \mid \lambda \geq 0 \right\}.$$

This is a direct sum (the Minkowski sum) of a linear subspace and a cone. To obtain its dual cone, the following lemma is useful.

Lemma. *Let K_1 be a cone, and let L be a linear subspace. Let $K = K_1 \oplus L$. Then, the dual cone of K is $K^* = K_1^* \cap L^\perp$.*

Because

$$K_1^* = \left\{ \lambda (h_{(0)} \otimes \psi_{(1)}) \mid \lambda \leq 0 \right\} \oplus \mathrm{span} \left\{ h_{(0)} \otimes \psi_{(1)} \right\}^\perp$$

and

$$L^\perp = \mathrm{span} \left\{ h_{(0)} \otimes \psi_{(j)},\ j = 1, \dots, p-1;\ h_{(i)} \otimes \psi_{(j)},\ i = 1, \dots, k-2,\ j = 1, \dots, p-1 \right\},$$

we have

$$N_{\phi(a,\theta)}(\mathrm{co}(M)) = K^* = K_1^* \cap L^\perp = \bigl\{\lambda(h_{(0)} \otimes \psi_{(1)}) \mid \lambda \le 0\bigr\} \oplus \mathrm{span}\bigl\{ h_{(0)} \otimes \psi_{(j)},\ j = 2, \dots, p-1; \ h_{(i)} \otimes \psi_{(j)},\ i = 1, \dots, k-2,\ j = 1, \dots, p-1\bigr\},$$

with $\dim N_{\phi(a,\theta)}(\mathrm{co}(M)) = pk - p - k + 1$.

The second fundamental form of $\mathrm{co}(\partial M)$ at $\phi(a,\theta)$ is

$$\left(\frac{\partial^2 \phi(a,\theta)}{\partial\theta_i \partial\theta_j}\right)^\top v = \left(\frac{\partial^2 h}{\partial\theta_i \partial\theta_j} \otimes \psi\right)^\top v, \quad v \in N_{\phi(a,\theta)}(\mathrm{co}(M)). \tag{A.1.7}$$

We can easily see that the second fundamental form (A.1.7) is always zero. Therefore, the contribution of the boundary to w'_{d-e} in (3.3.4) is only to case $e = 0$. That is, all w'_i but w'_{k-1} are zero. The contribution of the boundary $\{\phi(a,\theta) \mid \theta \in \Theta\}$ to w'_{k-1} is

$$\begin{aligned}\frac{1}{\Omega_{k-1-0}\Omega_{p(k-1)-(k-1)+0}} \int_{\partial M} |G|^{\frac{1}{2}} \mathrm{d}\theta \int_{N_{\phi(a,\theta)}(\mathrm{co}(M)) \cap \mathbb{S}^{n-1}} \mathrm{d}v =& \frac{\mathrm{Vol}(\mathbb{S}^{k-2})\mathrm{Vol}(\text{half of } \mathbb{S}^{pk-p-k})}{\Omega_{k-1}\Omega_{pk-p-k+1}} \\ =& \frac{1}{2}.\end{aligned}$$

The contribution of the other boundary $\{\phi(b,\theta) \mid \theta \in \Theta\}$ to w_{k-1} has the same value of $1/2$. Moreover, if the number of connected components of Γ exceeds one, we need to select all boundaries. Since the number of boundaries is $2\chi(\Gamma)$, we have

$$w'_{k-1} = \frac{1}{2} \times 2\chi(\Gamma) = \chi(\Gamma). \tag{A.1.8}$$

Substituting (A.1.6) and (A.1.8) into (3.3.2) yields (4.2.4).

A.2 Proof of Theorem 4.2.4

We apply the formula (3.3.5) for θ_c to the case where M is given in (4.2.1).

Let

$$u = \phi(\tilde{x}, \tilde{\theta}) = h(\tilde{\theta}) \otimes \psi(\tilde{x}), \quad v = \phi(x, \theta) = h(\theta) \otimes \psi(x),$$

and write $h = h(\theta)$, $\tilde{h} = h(\tilde{\theta})$, $\psi = \psi(x)$, $\tilde{\psi} = \psi(\tilde{x})$. We discuss the two cases (i) $\psi \in \mathrm{Int}\Gamma$ and (ii) $\psi \in \partial\Gamma$ separately.

Case (i). Suppose that $\psi(x) \in \mathrm{Int}\Gamma$. Write $\phi_{\theta_i} = (\partial h(\theta)/\partial\theta_i) \otimes \psi(x)$, $\phi_x = h(\theta) \otimes \psi_x(x)$, $\psi_x = \partial\psi(x)/\partial x$.

The orthogonal projection matrix onto the space $T_\phi(\mathrm{co}(M)) = \mathrm{span}\{\phi, \phi_{\theta_i}, \phi_x\}$ is

$$
\begin{aligned}
P_v &= \begin{pmatrix} \phi & \phi_{\theta_i} & \phi_x \end{pmatrix}_{p(k-1)\times k} \left(\begin{pmatrix} \phi^\top \\ \phi_{\theta_i}^\top \\ \phi_x^\top \end{pmatrix} \begin{pmatrix} \phi & \phi_{\theta_i} & \phi_x \end{pmatrix} \right)^{-1}_{k\times k} \begin{pmatrix} \phi^\top \\ \phi_{\theta_i}^\top \\ \phi_x^\top \end{pmatrix}_{k\times p(k-1)} \\
&= \begin{pmatrix} \phi & \phi_{\theta_i} & \phi_x \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & G(\theta) & 0 \\ 0 & 0 & g(x) \end{pmatrix}^{-1} \begin{pmatrix} \phi^\top \\ \phi_{\theta_i}^\top \\ \phi_x^\top \end{pmatrix} \\
&= \phi\phi^\top + (\phi_{\theta_i})G(\theta)^{-1}(\phi_{\theta_i})^\top + \frac{1}{g(x)}\phi_x\phi_x^\top \\
&= hh^\top \otimes \psi\psi^\top + (I_{k-1} - hh^\top) \otimes \psi\psi^\top + \frac{1}{g}hh^\top \otimes \psi_x\psi_x^\top \\
&= I_{k-1} \otimes \psi\psi^\top + \frac{1}{g}hh^\top \otimes \psi_x\psi_x^\top .
\end{aligned}
$$

As $P_v^\perp(w) = I(w) - P_v(w)$,

$$
\frac{(1-u^\top v)^2}{\|(I_n - P_v)u\|^2} = \frac{(1-(\tilde{h}\otimes\tilde{\psi})^\top(h\otimes\psi))^2}{\|(I_{k-1}\otimes I_p - I_{k-1}\otimes\psi\psi^\top - \frac{1}{g}hh^\top\otimes\psi_x\psi_x^\top)(\tilde{h}\otimes\tilde{\psi})\|^2}. \tag{A.2.1}
$$

Let $s = \psi^\top(x)\psi(\tilde{x}) = \psi^\top\tilde{\psi}$, $r = \psi_x^\top(x)\psi(\tilde{x}) = \psi_x^\top\tilde{\psi}$, $\alpha = h(\theta)^\top h(\tilde{\theta}) = h^\top\tilde{h}$. The numerator in (A.2.1) is $(1-\alpha s)^2$. The denominator in (A.2.1) is

$$
\begin{aligned}
\left\| \tilde{h}\otimes\tilde{\psi} - s\tilde{h}\otimes\psi - \frac{\alpha r}{g} h\otimes\psi_x \right\|^2 &= 1 + s^2 + \frac{\alpha^2 r^2}{g} - 2s^2 - 2\frac{\alpha r}{g}\alpha r \\
&= 1 - s^2 - \frac{\alpha^2 r^2}{g} = 1 - s^2 - \alpha^2 t^2,
\end{aligned}
$$

where $t = r/\sqrt{g} = \psi_x(x)^\top\psi(\tilde{x})/\|\psi_x(x)\|$. Hence,

$$
(\text{A.2.1}) = \frac{(1-\alpha s)^2}{1 - s^2 - \alpha^2 t^2}.
$$

The infimum is taken over "$x \neq \tilde{x}$ or $\theta \neq \tilde{\theta}$", or equivalently, "$x \neq \tilde{x}$ or $\alpha \neq 1$". However, when $x = \tilde{x}$ and $\alpha \neq 1$, the argument of the infimum is $(1-\alpha)^2/0 = \infty$. Therefore, we can exclude this case $x = \tilde{x}$ from the infimum argument.

Case (ii). Suppose that $\psi(x) \in \partial\Gamma$. Fix a point on the boundary

$$
v = \phi(x,\theta) = h(\theta)\otimes\psi(x) \in \partial M.
$$

The support cone of $\mathrm{co}(M)$ at v is

$$S_v(\mathrm{co}(M)) = \mathrm{span}\{\phi, \phi_{\theta_i}\} \oplus \{\lambda\varepsilon\phi_x \mid \lambda \geq 0\}$$

where $\varepsilon = \varepsilon(x) = 1$ if ψ_x is inward to Γ, $\varepsilon = -1$ if ψ_x is outward to Γ.

The orthogonal projection operator onto the cone $S_v(\mathrm{co}(M))$ is $w \mapsto P_v(w)$, where

$$\begin{aligned} P_v(w) =& \phi\phi^\top w + \phi_\theta G^{-1}(\theta)\phi_\theta^{-1} w + \frac{\phi_x}{\|\phi_x\|^2}\max\{0, \varepsilon\phi_x^\top w\} \\ =& (I_{k-1} \otimes \psi\psi^\top)w + \frac{\phi_x}{\|\phi_x\|^2}\max\{0, \varepsilon\phi_x^\top w\}. \end{aligned}$$

Hence,

$$P_v^\perp(w) = w - P_v(w) = w - (I_k \otimes \psi\psi^\top)w - \frac{h \otimes \psi_x}{g}\max\{0, \varepsilon(h \otimes \psi_x)w\}.$$

Substituting $u = \phi(\tilde{x}, \tilde{\theta}) = \tilde{h} \otimes \tilde{\psi}$,

$$P_v^\perp(u - v) = \tilde{h} \otimes \tilde{\psi} - s\tilde{h} \otimes \psi - \frac{\max\{0, \varepsilon\alpha r\}}{g} h \otimes \psi_x,$$

$$\begin{aligned} \|P_v^\perp(u - v)\|^2 =& 1 + s^2 + \frac{\max\{0, \varepsilon\alpha r\}^2}{g} - 2s^2 - 2\alpha r\frac{\max\{0, \varepsilon\alpha r\}}{g} \\ =& 1 - s^2 - \frac{\max\{0, \varepsilon\alpha r\}^2}{g} = 1 - s^2 - \max\{0, \varepsilon\alpha t\}^2, \end{aligned}$$

and we have

$$\frac{(1 - u^\top v)^2}{\|P_v^\perp(u - v)\|^2} = \frac{(1 - \alpha s)^2}{1 - s^2 - \max\{0, \varepsilon\alpha t\}^2}.$$

For the same reason as in case (i), the infimum is taken over the set $x \neq \tilde{x}$ and $\alpha \in [-1, 1]$.

A.3 Proof of Theorem 4.2.5

We use the same notations as in the proofs of Theorems 4.2.2 and 4.2.4. The local critical radius $\theta_{c,loc}$ defined by (3.3.6) is rewritten as

$$\tan^2\theta_{c,loc} = \liminf_{|x-\tilde{x}|\to 0,\,\alpha\to 1} \frac{(1-\alpha s)^2}{1-s^2-\alpha^2 t^2}.$$

Let $\tilde{x} = x + \Delta$ and $\alpha = 1-\delta$, and consider $\Delta \to 0$ and $\delta \to 0$. Write $\psi_x = \partial\psi/\partial x$, $\psi_{xx} = \partial^2\psi/\partial x^2$, etc., and $g = \psi_x^\top \psi_x$, $\gamma = \psi_{xx}^\top \psi_x$, and $\eta = \psi_{xx}^\top \psi_{xx}$ as before. Noting that

$$\begin{aligned}
&0 = \mathrm{d}(\psi^\top\psi)/\mathrm{d}x = 2\psi_x^\top\psi, \\
&0 = \mathrm{d}^2(\psi^\top\psi)/\mathrm{d}x^2 = 2\psi_{xx}^\top\psi + 2\psi_x^\top\psi_x, \\
&0 = \mathrm{d}^3(\psi^\top\psi)/\mathrm{d}x^3 = 2\psi_{xxx}^\top\psi + 6\psi_{xx}^\top\psi_x, \\
&0 = d^4(\psi^\top\psi)/dx^4 = 2\psi_{xxxx}^\top\psi + 8\psi_{xxx}^\top\psi_x + 6\psi_{xx}^\top\psi_{xx},
\end{aligned}$$

we have

$$\psi_{xx}^\top\psi = -g, \quad \psi_{xxx}^\top\psi = -3\gamma, \quad \psi_{xxxx}^\top\psi + 4\psi_{xxx}^\top\psi_x = -3\eta.$$

Substituting these, we have

$$\begin{aligned}
s = \psi(x)^\top\psi(\tilde{x}) =& \psi^\top\left(\psi + \psi_x\Delta + \frac{1}{2}\psi_{xx}\Delta^2 + \frac{1}{6}\psi_{xxx}\Delta^3 + \frac{1}{24}\psi_{xxxx}\Delta^4\right) + o(\Delta^4) \\
=& 1 - \frac{1}{2}g\Delta^2 - \frac{1}{2}\gamma\Delta^3 + \frac{1}{24}\psi_{xxxx}^\top\psi\Delta^4 + o(\Delta^4), \\
r = \psi_x(x)^\top\psi(\tilde{x}) =& \psi_x^\top\left(\psi + \psi_x\Delta + \frac{1}{2}\psi_{xx}\Delta^2 + \frac{1}{6}\psi_{xxx}\Delta^3\right) + o(\Delta^3) \\
=& g\Delta + \frac{1}{2}\gamma\Delta^2 + \frac{1}{6}\psi_{xxx}^\top\psi_x\Delta^3 + o(\Delta^3), \\
t^2 = \frac{r^2}{g} =& g\Delta^2 + \gamma\Delta^3 + \frac{1}{4g}\gamma^2\Delta^4 + \frac{1}{3}\psi_{xxx}^\top\psi_x\Delta^4 + o(\Delta^4),
\end{aligned}$$

and

$$\begin{aligned}1-s^2-t^2 =&(1-s)(2-(1-s))-t^2 = 2(1-s)-(1-s)^2-t^2\\ =&2\left(\frac{1}{2}g\Delta^2+\frac{1}{2}\gamma\Delta^3-\frac{1}{24}\psi_{xxxx}^\top\psi\Delta^4\right)-\left(\frac{1}{2}g\Delta^2\right)^2\\ &-\left(g\Delta^2+\gamma\Delta^3+\frac{1}{4g}\gamma^2\Delta^4+\frac{1}{3}\psi_{xxx}^\top\psi_x\Delta^4\right)+o(\Delta^4)\\ =&\frac{1}{4}\left(\eta-g^2-\frac{\gamma^2}{g}\right)\Delta^4+o(\Delta^4)=\frac{1}{4}\kappa g^2\Delta^4+o(\Delta^4),\end{aligned}$$

where

$$\kappa=\kappa(x)=\frac{\eta}{g^2}-\frac{\gamma^2}{g^3}-1.$$

Note that κ is nonnegative because

$$0\le\det\begin{pmatrix}\psi^\top\\ \psi_x^\top\\ \psi_{xx}\cdots\end{pmatrix}\begin{pmatrix}\psi & \psi_x & \psi_{xx}\end{pmatrix}=\det\begin{pmatrix}1&0&-g\\0&g&\gamma\\-g&\gamma&\eta\end{pmatrix}=\kappa g^3.$$

For the order of δ, we consider two cases: (i) $\delta/\Delta^2\sim gc$ ($0\le c<\infty$) and (ii) $\Delta^2/\delta\sim 0$.

For case (i), noting that $\alpha=1-\delta$, $1-s\sim\frac{1}{2}g\Delta^2$, $t^2\sim g\Delta^2$,

$$(1-\alpha s)^2=(1-s+\delta-\delta(1-s))^2\sim(1-s+\delta)^2\sim\left(\frac{1}{2}g\Delta^2+gc\Delta^2\right)^2=\frac{1}{4}g^2(1+2c)^2\Delta^2,$$

$$1-s^2-\alpha^2t^2=1-s^2-t^2+2\delta t^2-\delta^2t^2\sim\frac{1}{4}g^2(\kappa+8c)\Delta^4,$$

and hence

$$\frac{(1-\alpha s)^2}{1-s^2-\alpha^2t^2}\sim\frac{(1+2c)^2}{\kappa+8c}. \tag{A.3.1}$$

We consider the minimum value of (A.3.1) for $c\ge 0$. For $c>-\kappa/8$, (A.3.1) has a unique minimum value $1-\kappa/4$ at $c=(2-\kappa)/4$. Therefore,

$$\min_{c\ge0}\frac{(1+2c)^2}{\kappa+8c}=\begin{cases}1-\dfrac{\kappa}{4} & (\kappa\le 2),\\ \dfrac{1}{\kappa} & (\kappa\ge 2).\end{cases}$$

For case (ii),

$$(1-\alpha s)^2 = (1-s+\delta-\delta(1-s))^2 \sim \delta^2,$$
$$1-s^2-\alpha^2 t^2 = 1-s^2-t^2+2\delta t^2-\delta^2 t^2 \sim 2g\Delta^2\delta,$$

and

$$\frac{(1-\alpha s)^2}{1-s^2-\alpha^2 t^2} \sim \frac{\delta}{2g\Delta^2} \to \infty.$$

In summary, we have

$$\tan^2\theta_{c,loc} = \min\left\{\inf_{x:\kappa(x)\le 2}\left(1-\frac{\kappa(x)}{4}\right), \inf_{x:\kappa(x)\ge 2}\frac{1}{\kappa(x)}\right\}.$$

A.4 Coverage probability under model misspecification

First, note that the best parameter under the model $\beta_i^\top f(x)$ is given by $\beta_i^* = \Sigma X^\top g_i$, where

$$X = \begin{pmatrix} f(x_1)^\top \\ \vdots \\ f(x_n)^\top \end{pmatrix}, \quad \Sigma = (X^\top X)^{-1} = \left(\sum_{i=1}^n f(x_i)f(x_i)^\top\right)^{-1}, \quad g_i = \begin{pmatrix} g_i(x_1) \\ \vdots \\ g_i(x_n) \end{pmatrix}.$$

Let

$$y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in} \end{pmatrix}, \quad \varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{i1} \end{pmatrix}.$$

The least square estimator for β_i^* is $\widehat{\beta}_i^* = \Sigma X^\top y_i$, which is distributed as $\mathcal{N}_p(\beta_i^*, \Sigma)$. Let $b_{1-\alpha}$ be the threshold for $1-\alpha$ bands when the assumed model is the true model. The approximate value of $b_{1-\alpha}$ can be obtained by the tube method.

On the other hand, when the true model is $g_i(x)$, the coverage probability becomes

$$\begin{aligned}
&\Pr\left\{\left|\sum_{i=1}^{k} c_i(\widehat{\beta}_i^*)^\top f(x) - \sum_{i=1}^{k} c_i g_i(x)\right| \le b_{1-\alpha}\sqrt{f(x)^\top \Sigma f(x)} \ \text{ for all } x \in \mathcal{X},\ c \in \mathcal{C}\right\} \\
&= \Pr\left\{\max_{x\in\mathcal{X},\, c\in\mathcal{C}} \frac{\sum_{i=1}^{k} c_i(\widehat{\beta}_i^*)^\top f(x) - \sum_{i=1}^{k} c_i g_i(x)}{\sqrt{f(x)^\top \Sigma f(x)}} \le b_{1-\alpha}\right\} \\
&= \Pr\left(\max_{x\in\mathcal{X},\, c\in\mathcal{C}} \left[\frac{\sum_{i=1}^{k} c_i(\widehat{\beta}_i^* - \beta_i^*)^\top f(x)}{\sqrt{f(x)^\top \Sigma f(x)}} + \frac{\sum_{i=1}^{k} c_i\{(\beta_i^*)^\top f(x) - g_i(x)\}}{\sqrt{f(x)^\top \Sigma f(x)}}\right] \le b_{1-\alpha}\right).
\end{aligned} \tag{A.4.1}$$

Noting that, for the two functions $h_1(y)$ and $h_2(y)$ on $\mathcal{Y}$, if $\max_{y\in\mathcal{Y}}(-h_i(y)) = \max_{y\in\mathcal{Y}} h_i(y)$, then

$$\max_y h_1(y) \le \max_y(h_1(y)+h_2(y)) + \max_y(-h_2(y)) = \max_y(h_1(y)+h_2(y)) + \max_y h_2(y),$$

hence,

$$\max_y h_1(y) - \max_y h_2(y) \le \max_y(h_1(y)+h_2(y)) \le \max_y h_1(y) + \max_y h_2(y).$$

Therefore, (A.4.1) is bounded below and above by $1-\overline{P}(b_{1-\alpha}-\delta)$ and $1-\overline{P}(b_{1-\alpha}+\delta)$, respectively, where

$$\begin{aligned}
\delta &= \max_{x\in\mathcal{X},\, c\in\mathcal{C}} \frac{\sum_{i=1}^{k} c_i\{(\beta_i^*)^\top f(x) - g_i(x)\}}{\sqrt{f(x)^\top \Sigma f(x)}} \\
&= \max_{x\in\mathcal{X}} \sqrt{\frac{\sum_{i=1}^{k}[(\beta_i^*)^\top f(x) - g_i(x) - \frac{1}{k}\sum_{i=1}^{k}\{(\beta_i^*)^\top f(x) - g_i(x)\}]^2}{f(x)^\top \Sigma f(x)}}
\end{aligned}$$

as given in (4.4.2), and

$$\overline{P}(b) = \Pr\left\{\max_{x\in\mathcal{X},\, c\in\mathcal{C}} \frac{\sum_{i=1}^{k} c_i(\widehat{\beta}_i^* - \beta_i^*)^\top f(x)}{\sqrt{f(x)^\top \Sigma f(x)}} \ge b\right\}.$$

An upper bound for the bias of coverage probability for a $1-\alpha$ confidence band is

$$\max\left\{\alpha - \overline{P}(b_{1-\alpha}+\delta), \overline{P}(b_{1-\alpha}-\delta) - \alpha\right\},$$

which is approximated by

$$\Delta = \max\bigl\{\alpha - \overline{P}_{\mathrm{tube}}\bigl(b_{\mathrm{tube},1-\alpha} + \delta\bigr), \overline{P}_{\mathrm{tube}}\bigl(b_{\mathrm{tube},1-\alpha} - \delta\bigr) - \alpha\bigr\}$$

in (4.4.3), where $\overline{P}_{\mathrm{tube}}(b)$ is the tube approximation formula for $P(b)$ given in (4.2.4), and $b_{\mathrm{tube},1-\alpha}$ is the solution of $\overline{P}_{\mathrm{tube}}(b) = \alpha$.

Note that (A.4.1) is

$$\Pr\left(\max_{x\in\mathcal{X}} \frac{\sum_{i=1}^{k}[(\widehat{\beta}_i^*)^\top f(x) - g_i(x) - \frac{1}{k}\sum_{i=1}^{k}\{(\widehat{\beta}_i^*)^\top f(x) - g_i(x)\}]^2}{f(x)^\top \Sigma f(x)} \le b_{1-\alpha}^2\right),$$

which is used for the simulation study.

References

Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*, Springer.

Anderson, T.W. (2009). *An Introduction to Multivariate Statistical Analysis*, Wiley.

de Boor, C. (1978). *A Practical Guide to Splines*, Springer.

Casella, G. and Hwang, J. T. G. (2012). Shrinkage confidence procedures, *Statistical Science*, **27** (1), 51–60.

Faraway, J. J. and Sun, J. (1995). Simultaneous confidence bands for linear regression with heteroscedastic errors, *Journal of the American Statistical Association*, **90** (431), 1094–1098.

Federer, H. (1959). Curvature measures, *Transactions of the American Mathematical Society*, **93** (3), 418–491.

Friston, K.J., Frith, C.D., Liddle, P.F., Dolan, R.J., Lammertsma, A.A. and Frackowiak. R.S.J. (1990). The relationship between global and local changes in PET scans, *Journal of Cerebral Blood Flow & Metabolism*, **10**, 458–466.

Friston, K.J., Frith, C.D., Passingham, R.E., Dolan, R.J., Liddle, P.F. and Frackowiak. R.S.J. (1992). Entropy and cortical activity: Information theory and PET findings, *Cerebral Cortex*, **2**, 259–267.

Friston, K.J., Worsley, K.J., Frackowiak. R.S.J., Mazziotta, J.C., and Evans, A.C. (1994). Assessing the significance of focal activations using their spatial extent, *Human Brain Mapping*, **1**, 210–220.

Hotelling, H. (1939). Tubes and spheres in n-spaces, and a class of statistical problems, *American Journal of Mathematics*, **61** (2), 440–460.

Hotelling, H. (1951). A generalized T-Test and measure of multivariate dispersion, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probablity*, 23–41.

Jamshidian, M., Liu, W., and Bretz, F. (2010). Simultaneous confidence bands for all contrasts of three or more simple linear regression models over an interval, *Computational Statistics and Data Analysis*, **54** (6), 1475–1483.

Johansen, S. and Johnstone, I. M. (1990). Hotelling's theorem on the volume of tubes: Some illustrations in simultaneous inference and data analysis, *The Annals of Statistics*, **18** (2), 652–684.

Johnstone, I. and Siegmund, D. (1989). On Hotelling's formula for the volume of tubes and Naiman's inequality, *The Annals of Statistics*, **17** (1), 184–194.

Krivobokova, T., Kneib, T., and Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators, *Journal of the American Statistical Association*, **105** (490), 852–863.

Kuriki, S. and Takemura, A. (2001). Tail probabilities of the maxima of multilinear forms and their applications, *The Annals of Statistics*, **29** (2), 328–371.

Kuriki, S. and Takemura, A. (2009). Volume of tubes and the distribution of the maximum of a Gaussian random field, Selected Papers on Probability and Statistics, American Mathematical Society Translations Series 2, **227** (2), 25–48.

Leeb, H., Pötscher, B. M., and Ewald, K. (2015). On various confidence intervals post-model-selection, *Statistical Science*, **30** (2), 216–227.

Liu, W. (2010). *Simultaneous Inference in Regression*, Chapman & Hall/CRC.

Liu, W., Jamshidian, M., and Zhang, Y. (2004). Multiple comparison of several linear regression models, *Journal of the American Statistical Association*, **99** (466), 395–403.

Liu, W., Lin, S., and Piegorsch, W. W. (2008). Construction of exact simultaneous confidence bands for a simple linear regression model, *International Statistical Review*, **76** (1), 39–57.

Liu, W., Wynn, H. P., and Hayter, A. J. (2008). Statistical inferences for linear regression models when the covariates have functional relationships: polynomial regression, *Journal of Statistical Computation and Simulation*, **78** (4), 315–324.

Lu, X. and Chen, J. T. (2009). Exact simultaneous confidence segments for all contrast comparisons, *Journal of Statistical Planning and Inference*, **139** (8), 2816–2822.

Lu, X. and Kuriki, S. (2017) Simultaneous confidence bands for contrasts between several nonlinear regression curves. *Journal of Multivariate Analysis*, **155**, 83–104.

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*, 2nd ed., Wiley.

Naiman, D. Q. (1986). Conservative confidence bands in curvilinear regression, *The Annals of Statistics*, **14** (3), 896–906.

Scheffé, H. (1953). A method for judging all contrasts in analysis of variance. *Biometrika*, **40** (1-2), 87–104.

Scheffé, H. (1959). *The Annals of Variance*, Wiley.

Spurrier, J. D. (1999). Exact confidence bounds for all contrasts of three or more regression lines, *Journal of the American Statistical Association*, **94** (446), 483–488.

Spurrier, J. D. (2002). Exact multiple comparisons of three or more regression lines: Pairwise comparisons and comparisons with a control, *Biometrical Journal*, **44** (7), 801–812.

Sun, J. (1993). Tail probabilities of the maxima of Gaussian random fields, *The Annals of Probability*, **21** (1), 34–71.

Sun, J., and Loader, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing, *The Annals of Statistics*, **22** (3), 1328–1346.

Sun, J., Loader, C., and McCormick, W. P. (2000). Confidence bands in generalized linear models, *The Annals of Statistics*, **28** (2), 429–460.

Sun, J., Raz, J., and Faraway, J. J. (1999). Confidence bands for growth and response curves, *Statistica Sinica*, **9**, 679–698.

Takada, T., Mita, A., Maeno, A., Sakai, T., Shitara, H., Kikkawa, Y., Moriwaki, K, Yonekawa, H., and Shiroishi, T. (2008). Mouse inter-subspecific consomic strains for genetic dissection of quantitative complex traits, *Genome Research*, **18** (3), 500–508.
`http://genome.cshlp.org/content/18/3/500`

Takemura, A. and Kuriki, S. (2002). On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains, *The Annals of Applied Probability*, **12** (2), 768–796.

Uusipaikka, E. (1983). Exact confidence bands for linear regression over intervals, *Journal of the American Statistical Association*, **78** (383), 638–644.

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

Wolfram Research, Inc. (2016). Mathematica, Version 11.0.

Working, H. and Hotelling, H. (1929). Applications of the theory of error to the interpretation of trends, *Journal of the American Statistical Association*, **24** (165), 73–85.

Worsley, K. (1995). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics, *Advances in Applied Probability*, **27** (4), 943–959.

Wynn, H. P. and Bloomfield, P. (1971). Simultaneous confidence bands in regression analysis, *Journal of the Royal Statistical Society, Series B*, **33** (2), 202–217.

Zhou, S., Wolfe, D. A., and Shen, X. (1998). Local asymptotics for regression splines and confidence regions, *The Annals of Statistics*, **26** (5), 1760–1782.