UNIVERSITY OF SOUTHAMPTON

Simultaneous Confidence Bands For Linear And Logistic Regression Models

by

Shan Lin

A thesis submitted in partial fulfillment for the degree of
Doctor of Philosophy

in the

Faculty of Engineering, Science and Mathematics
School of Mathematics

July 2007

UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF MATHEMATICS

Doctor of Philosophy

SIMULTANEOUS CONFIDENCE BANDS FOR LINEAR AND LOGISTIC REGRESSION MODELS

by

Shan Lin

ABSTRACT

This thesis considers the construction of simultaneous confidence bands for a normal-error linear regression model and a linear logistic regression model with a binary response variable respectively. For linear regression, three general methods are summarized to construct exact one-sided and two-sided confidence bands over an ellipsoidal restricted region of the predictor space, and they are found to have the equivalent formulae for calculating critical values. Also, several methods are available to construct confidence bands over a rectangular region. We compare these methods in terms of the critical value. For logistic regression, several methods are considered for the construction of confidence bands with or without predictor constraint, which is based on the asymptotic normality of the estimator. Simulation studies are provided to assess the performances of some key bands. Several useful conclusions can be drawn.

Contents

List of Figures

List of Tables

Acknowledgements

I wish to thank Professor Wei Liu very much for providing many valuable ideas and kind supervision, from which I have greatly benefited during my PhD period. Also, the staff in School of Mathematics gave me lots of timely help which drove me to move forward. I would like to thank them as well. Moreover, I got much concern and encouragement from my wife and our parents, who made me feel and enjoy the wonderful life outside the absolutely boring studies. So this thesis is to them, and additionally to my unborn daughter - Peiyao Lin.

Shan Lin
University of Southampton
July 2007

Chapter 1

Introduction

1.1 Background

Consider the classical normal-error linear regression model

$$\mathbf{Y} = X\beta + \varepsilon$$

where $\mathbf{Y}_{n\times 1}$ is the vector of the observed responses, $X_{n\times p}$ is the design matrix with the first column given by $(1,\ldots,1)^T$ and the jth $(2 \le j \le p)$ column given by $(x_{1,j},\ldots,x_{n,j})^T$, $\beta = (\beta_1,\ldots,\beta_p)^T$ is the vector of regression coefficients, and ε is the error vector which has the $N_n(\mathbf{0},\sigma^2 I)$ distribution with σ^2 unknown. Assume $X^T X$ is non-singular, then the least squares estimator of β is given by $\hat{\beta} = (X^T X)^{-1} X^T \mathbf{Y}$ which has the $N_p(\beta, \sigma^2 (X^T X)^{-1})$ distribution. Let $\hat{\sigma}^2$ denote the usual unbiased estimator of σ^2, then $\hat{\sigma}^2 \sim \sigma^2 \chi^2_\nu/\nu$ with the degree of freedom $\nu = n - p$ and is independent of $\hat{\beta}$.

For statistical inference, the commonly considered pointwise confidence interval plays an important role which is concerned for the mean response $\mathbf{x}_0^T\beta$ at one specific point $\mathbf{x}_0$. It has the form given by

$$\mathbf{x}_0^T\beta \in \mathbf{x}_0^T\hat{\beta} \pm t_{\nu,1-\alpha/2}\hat{\sigma}\sqrt{\mathbf{x}_0^T(X^T X)^{-1}\mathbf{x}_0}, \tag{1.1}$$

where $t_{\nu,1-\alpha/2}$ is the percentage point of a t random variable with ν degrees of freedom that leaves a probability $\alpha/2$ in the upper tail, and so $1-\alpha/2$ in the lower tail.

A simultaneous confidence band is constructed for the mean responses $\mathbf{x}^T\boldsymbol{\beta}$ for all possible values of $\mathbf{x}$ within a given region $\mathcal{X}$ of $p-1$ predictor variables. The most popular simultaneous confidence band is of hyperbolic shape, and has the following form

$$\mathbf{x}^T\boldsymbol{\beta} \in \mathbf{x}^T\hat{\boldsymbol{\beta}} \pm c\hat{\sigma}\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}} \quad \text{for all } \mathbf{x} \in \mathcal{X}, \tag{1.2}$$

where c is the critical value such that the confidence band has the simultaneous coverage probability equal to a preassigned confidence level $1-\alpha$. The key of constructing a confidence band is to find the appropriate critical value c. Another frequently mentioned confidence band is of fixed band width, which has the form given by

$$\mathbf{x}^T\boldsymbol{\beta} \in \mathbf{x}^T\hat{\boldsymbol{\beta}} \pm c\hat{\sigma} \quad \text{for all } \mathbf{x} \in \mathcal{X}. \tag{1.3}$$

It is of natural interest to compare the simultaneous confidence band with the pointwise confidence interval. The key difference between them is that the simultaneous confidence band is constructed for all possible $\mathbf{x}$ while the pointwise confidence interval is only at a specific point $\mathbf{x}_0$.

On the other hand, consider a confidence interval for the parameter vector $\boldsymbol{\beta}$, which is given by

$$\boldsymbol{\beta} \in \hat{\boldsymbol{\beta}} \pm t_{\nu,1-\alpha/2}\text{s.e.}(\hat{\boldsymbol{\beta}}), \tag{1.4}$$

where s.e.$(\hat{\boldsymbol{\beta}})$ is the standard error of $\hat{\boldsymbol{\beta}}$ and is formed by the square roots of the diagonal terms of the matrix $\hat{\sigma}^2(X^TX)^{-1}$. The confidence interval for $\boldsymbol{\beta}$ contains p individual confidence intervals for p regression coefficients respectively. And these individual intervals can be used to define a rectangular region in the parameter space. Note that this rectangular region is not a proper simultaneous confidence region for $\boldsymbol{\beta}$.

To obtain a simultaneous confidence region for $\boldsymbol{\beta}$, we start with the fact that $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(X^TX)^{-1})$. Define a $p \times p$ non-singular matrix P such that

$P^T P = (X^T X)^{-1}$. Then we have

$$\begin{aligned} & (P^{-1})^T(\hat{\beta} - \beta)/\sigma \sim N_p(0, I) \\ \Rightarrow \ & (\hat{\beta} - \beta)^T (P^T P)^{-1} (\hat{\beta} - \beta)/\sigma^2 \sim \chi_p^2 \\ \Rightarrow \ & \frac{(\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta)/p\sigma^2}{\hat{\sigma}^2/\sigma^2} \sim F_{p,\nu}, \end{aligned} \tag{1.5}$$

where χ_p^2 and $F_{p,\nu}$ denote the Chi-square distribution with p degrees of freedom and the F distribution with p and ν degrees of freedom. Therefore, a $(1-\alpha)$-level simultaneous confidence region for β can be obtained from the inequality

$$(\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta}) \leq p\hat{\sigma}^2 F_{p,\nu,1-\alpha}, \tag{1.6}$$

where $F_{p,\nu,1-\alpha}$ is the upper α point of the $F_{p,\nu}$ distribution. The equality obtained by changing "$\leq$" to "$=$" in (1.6) specifies the boundary of an ellipsoidal contour in the parameter space.

Note that the simultaneous confidence region for β in (1.6) can also be obtained from the simultaneous confidence band for $\mathbf{x}^T\beta$ in (1.2) when $\mathcal{X} = \mathcal{R}^{p-1}$ which is the setting in Scheffé (1953). Assume the band (1.2) has $1-\alpha$ confidence level. Then we have

$$\mathrm{P}\{\mathbf{x}^T\beta \in \mathbf{x}^T\hat{\beta} \pm c\hat{\sigma}\sqrt{\mathbf{x}^T(X^T X)^{-1}\mathbf{x}}, \mathbf{x} \in \mathcal{R}^{p-1}\} = 1 - \alpha. \tag{1.7}$$

With P consistently defined, we have the probability on the left-hand side of (1.7) further equal to

$$\begin{aligned} & \mathrm{P}\{\sup_{\mathbf{x}\in\mathcal{R}^{p-1}} |\mathbf{x}^T(\beta - \hat{\beta})| \leq c\hat{\sigma}\sqrt{\mathbf{x}^T(X^T X)^{-1}\mathbf{x}}\} \\ = \ & \mathrm{P}\{\sup_{\mathbf{x}\in\mathcal{R}^{p-1}} |\mathbf{x}^T P^T (P^{-1})^T(\beta - \hat{\beta})| \leq c\hat{\sigma}\sqrt{\mathbf{x}^T P^T P\mathbf{x}}\} \\ = \ & \mathrm{P}\{\sup_{\mathbf{x}\in\mathcal{R}^{p-1}} |(P\mathbf{x})^T \cdot (P^{-1})^T(\beta - \hat{\beta})| \leq c\hat{\sigma}\sqrt{(P\mathbf{x})^T(P\mathbf{x})}\} \\ = \ & \mathrm{P}\{\|P\mathbf{x}\| \cdot \|(P^{-1})^T(\beta - \hat{\beta})\| \leq c\hat{\sigma}\|P\mathbf{x}\|\} \\ = \ & \mathrm{P}\{[(P^{-1})^T(\beta - \hat{\beta})]^T[(P^{-1})^T(\beta - \hat{\beta})] \leq c^2\hat{\sigma}^2\} \\ = \ & \mathrm{P}\{(\beta - \hat{\beta})^T(P^T P)^{-1}(\beta - \hat{\beta}) \leq c^2\hat{\sigma}^2\} \\ = \ & \mathrm{P}\{(\beta - \hat{\beta})^T X^T X(\beta - \hat{\beta}) \leq c^2\hat{\sigma}^2\} = 1 - \alpha. \end{aligned} \tag{1.8}$$

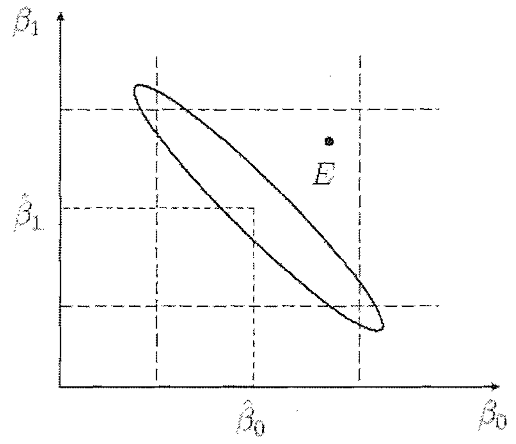

Figure 1.1: Simultaneous and individual confidence statements.

Therefore $c^2 = pF_{p,\nu,1-\alpha}$, and the link between (1.2) and (1.6) is obtained. For all $\mathbf{x} \in \mathcal{R}^{p-1}$, any point within the simultaneous confidence region for $\boldsymbol{\beta}$ in (1.6) one-to-one corresponds to a straight line which is completely inside the simultaneous confidence band for $\mathbf{x}^T\boldsymbol{\beta}$ in (1.2).

Figure 1.1 indicates a possible situation that may arise when $p = 2$. The $(1-\alpha)$-level simultaneous confidence region for (β_0, β_1) is displayed by the shin ellipse which encloses points of (β_0, β_1) that are considered as simultaneously appropriate for the true parameters. The individual $(1-\alpha)$-level confidence intervals for β_0 and β_1 specify the ranges for the candidates of the true parameters separately irrespective with the value of the other parameter. Both ellipse and the rectangular region are centered at the point of the estimates of the two parameters $(\hat{\beta}_0, \hat{\beta}_1)$. Note that a point, for example, E lying inside the rectangular region but outside the ellipse illustrates that the coordinates of the point E are regarded as reasonable for parameters β_0 and β_1 by the individual confidence intervals but not so by the simultaneous confidence region. For details, see, e.g., Draper and Smith (1998, pages 142-146). This thesis focuses on the construction of hyperbolic-shape simultaneous confidence bands rather than bands of other shapes or pointwise confidence intervals.

i	x_i	y_i
1	1.9	0.7
2	0.8	-1.0
3	1.1	-0.2
4	0.1	-1.2
5	-0.1	-0.1
6	4.4	3.4
7	4.6	0.0
8	1.6	0.8
9	5.5	3.7
10	3.4	2.0

Table 1.1: Observations for simple linear regression model

Next, we come to see two examples of constructing simultaneous confidence bands for a linear regression model. The first example is of one dimension, where we have 10 observations for the only predictor variable x and the response y respectively. These observations are given in Table 1.1. We fit this data using a simple linear regression model and construct a simultaneous confidence band over the restricted interval [-0.1, 5.5] with 95% confidence level. The critical value of the confidence band is 2.9201 compared with the critical value 2.3060 for the 95%-level pointwise confidence interval. The confidence band constructed is shown in Figure 1.2.

The second example is for two-dimensional case. Consider the acetylene data of Snee (1977) which was very popular in published papers and can be fitted by a bivariate linear regression model. We construct the 95%-level simultaneous confidence band over $\mathcal{X} = [1100, 1300] \times [5.3, 23]$ and then picture it in Figure 1.3. The critical value of the confidence band is 3.1137 while that of the pointwise confidence interval is 2.1604.

A simultaneous confidence band provides useful information on whereabouts of the true regression function. Any regression function which lies completely inside the confidence band over the whole given region of the predictor variables is deemed by the band as a plausible candidate of the true function; any regression function that lies outside the confidence band for at least one point in the given region of the predictor space is not considered as

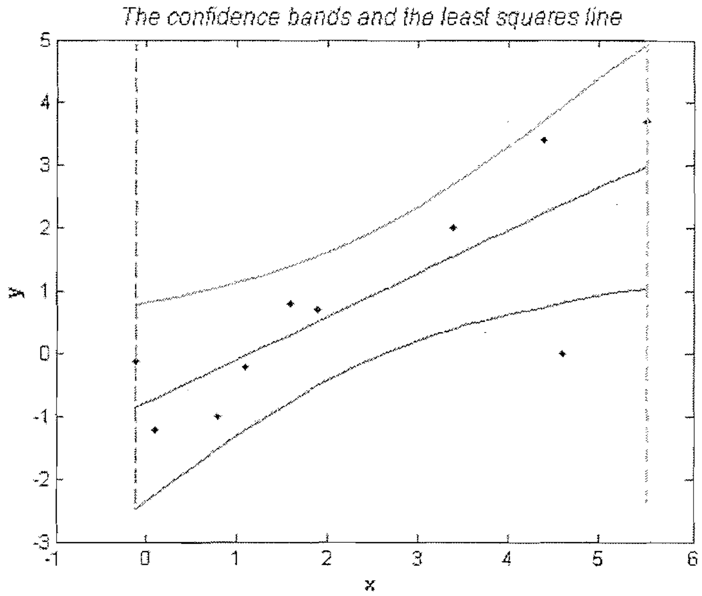

Figure 1.2: Confidence band for a simple linear regression

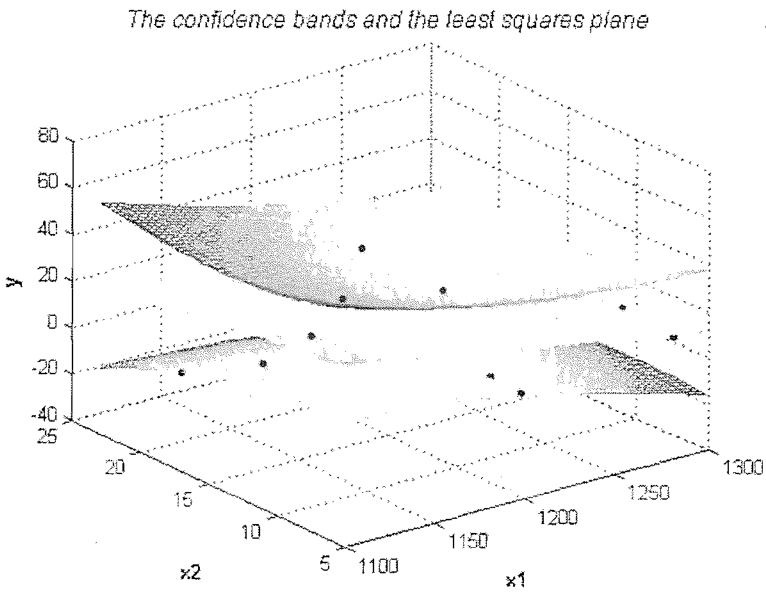

Figure 1.3: Confidence band for a bivariate linear regression

a potential candidate of the true function.

Specifically, a simultaneous confidence band can be used to test the following hypotheses

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \qquad \text{against} \qquad H_a : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$$

in the following way

$$\begin{aligned} &\text{reject} \quad H_0 \quad \text{if and only if} \\ &\qquad \mathbf{x}^T \boldsymbol{\beta}_0 \quad \text{is outside the band for at least one } \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{1.9}$$

This test is of size α since the confidence band has a simultaneous confidence level $1 - \alpha$.

The problem of constructing simultaneous confidence bands has a history going back to Working and Hotelling (1929). Scheffé (1953) considered the whole predictor space as the given region of predictor variables which is equivalent to setting no constraint at all on predictor variables.

For $p = 2$, that is, there is only one predictor variable, Gafarian (1964) considered a two-sided constant-width confidence band with the only predictor variable restricted in an interval. His effort was followed by Bowden (1970) who considered two-sided confidence bands of other shapes by making use of Hölder's inequality. Piegorsch *et al.* (2000) considered the calculation of the critical values of a family of confidence bands from Bowden (1970). Wynn and Bloomfield (1971) and Uusipaikka (1983) provided exact two-sided hyperbolic-shape confidence bands, with the band width proportional to the standard error of the estimated regression function, when the only predictor variable is restricted in an interval or the union of disjoint intervals. Bohrer and Francis (1972) proposed exact one-sided confidence bands with the only predictor variable constrained to an interval.

For $p > 2$, there are at least two predictor variables in the model. In such a case, the $(p-1)$-dimensional region $\mathcal{X}$ may have various forms. Construction of exact confidence bands becomes much harder. Bohrer (1967) considered a hyperbolic-shape confidence band when the predictor variables are all non-negative. Bohrer (1973) presented the construction of an exact one-sided confidence band over an ellipsoidal predictor region by evaluating a multivariate t probability. Halperin and Gurian (1968) provided conservative confidence bands over an ellipsoidal region. Wynn (1975) developed a general result on the calculation of the confidence levels for one-sided confidence bands in regression analysis. Casella and Strawderman (1980) proposed exact confidence bands over a region of the same shape. The most frequently used region is of rectangular shape, and it is given by

$$\mathcal{X}_R = \{(x_2, \ldots, x_p) : a_i \leq x_i \leq b_i, i = 2, \ldots, p\},$$

where $-\infty \leq a_i < b_i \leq \infty, i = 2, \ldots, p$ are given constants. Knafl, Sacks and Ylvisaker (1985) obtained an approximate two-sided hyperbolic-shape confidence band when $p \leq 3$ by using an up-crossing inequality. This approach was further developed in Faraway and Sun (1995), Sun and Loader (1994), and Sun, Loader and McCormick (2000) to produce approximate two-sided confidence bands for a more general regression model. However, multiple integrations are involved in the calculation of these approximations and the dimensionality of the integrations increases with p. Naiman (1986) discussed the construction of conservative simultaneous confidence bands for curvilinear regression functions by applying the tube volume theory. For the construction of confidence bands for a more general regression model, more references can be found in Johnstone and Siegmund (1989), Knowles and Siegmund (1989), Johansen and Johnstone (1990), and Sun, Loader and McCormick (2000). Recently, Liu, Jamshidian, Zhang and Donnelly (2005) proposed the simulation-based two-sided simultaneous confidence bands over a rectangular predictor space for generally $p > 2$, and the critical value based on this method can be as accurate as one expects if the number of simulations is set to be sufficiently large. Moreover, this simulation-based method can be adapted to the construction of one-sided confidence bands over a similar region. Liu, Jamshidian, Zhang and Bretz (2004) considered constructing two-sided constant-width confidence bands for a multiple regression model over a rectangular region by using both numerical integration and simulation.

The existing literatures of the construction of simultaneous confidence bands for logistic regression models are very limited. The main contributions to this area are: Brand, Pinnock and Jackson (1973) which described a method of obtaining a confidence band for a simple logistic regression based on the large sample distribution of the maximum likelihood estimators, Hauck (1983) which further developed the previous work to the multiple case by applying the Cauchy-Schwartz inequality, Piegorsch and Casella (1988) which first discussed the confidence bands for a logistic regression with restricted predictor variables, and Sun, Loader and McCormick (2000) which developed their approximate method of Sun and Loader (1994) applicable to the generalized linear models.

1.2 The organization of this thesis

We continue to introduce some concepts and basic tools in the rest of this chapter on large sample theory, which include some important inequalities and theorems required in the subsequent chapters but without explicit proof here. In Chapter 2, we describe the generalized linear models, specially, the logistic regression model, involving the large sample asymptotic distribution of the estimators and related inferences. In Chapter 3, our attention is focused on the construction of exact one-sided and two-sided hyperbolic-shape simultaneous confidence bands for a simple linear regression model with restricted predictor variable based on three methods. Chapter 4 continues to talk about the construction of confidence bands using the same methods for a multiple linear regression over an ellipsoidal region. In Chapter 5, we consider the construction of simultaneous confidence bands for a regression model over a rectangular region based on several methods and then compare these methods in terms of critical values. In Chapter 6, we discuss the construction of simultaneous confidence bands for a logistic regression model and then give simulation studies to check the goodness of the considered bands. Finally, Chapter 7 provides some main conclusions and the future work.

1.3 Concepts and basic tools

Definition 1.3.1 (Convergence in probability) *A sequence $\{T_n\}$ of random variables is said to converge in probability to a (possibly degenerate) random variable T, if for every positive numbers ε and η, there exists a positive integer $n_0 = n_0(\varepsilon, \eta)$, such that*

$$\mathrm{P}\{d(T_n, T) > \varepsilon\} < \eta, \quad n \geq n_0, \tag{1.10}$$

where $d(\cdot)$ denotes a distance function (or norm). This mode of convergence is usually expressed by $T_n - T \xrightarrow{P} 0$. In the case where T is non-stochastic, we may write $T_n \xrightarrow{P} T$.

Definition 1.3.2 (Convergence in distribution) *A sequence $\{T_n\}$ of random variables with distribution functions F_n is said to converge in distribution (or in law) to a (possibly degenerate) random variable T with a distribution function F, if for every $\varepsilon > 0$, there exists an integer $n_0 = n_0(\varepsilon)$, such that at every point of continuity of F,*

$$|F_n(x) - F(x)| < \varepsilon, \quad n \geq n_0. \tag{1.11}$$

This mode of convergence is denoted in this thesis by $T_n \xrightarrow{\mathcal{D}} T$.

Definition 1.3.3 (Almost sure convergence) *A sequence $\{T_n\}$ of random variables is said to converge almost surely (a.s.) to a (possibly degenerate) random variable T, if for every positive ε and η, there exists a positive integer $n_0 = n_0(\varepsilon, \eta)$, such that*

$$\mathrm{P}\{d(T_N, T) > \varepsilon \quad \text{for some } N \geq n\} < \eta, \quad n \geq n_0. \tag{1.12}$$

In symbols, we write this as $T_n - T \xrightarrow{a.s.} 0$, and if T is non-stochastic, it may also be written as $T_n \xrightarrow{a.s.} T$.

Theorem 1.3.1 (Chebyshev Inequality) *Let U be a non-negative random variable with a finite mean $\mu = \mathrm{E}(U)$. Then for every $t > 0$,*

$$\mathrm{P}\{U > t\mu\} \leq t^{-1}. \tag{1.13}$$

Theorem 1.3.2 (Lindeberg-Feller) *Let $X_k, k \geq 1$, be independent random variables such that* $\mathrm{E}(X_k) = \mu_k$ *and* $\mathrm{Var}(X_k) = \sigma_k^2, k \geq 1$*; also let* $T_n = \sum_{k=1}^n X_k$, $\xi_n = \mathrm{E}(T_n) = \sum_{k=1}^n \mu_k$, $s_n^2 = \mathrm{Var}(T_n) = \sum_{k=1}^n \sigma_k^2$ *and* $Z_n = (T_n - \xi_n)/s_n = \sum_{k=1}^n Y_{nk}$ *where* $Y_{nk} = (X_k - \mu_k)/s_n$. *Consider the following conditions:*

A) Uniform asymptotic negligibility condition:

$$\max_{1\leq k\leq n} \frac{\sigma_k^2}{s_n^2} \to 0 \quad as \quad n \to \infty.$$

B) Asymptotic normality condition:

$$\mathrm{P}\{Z_n \leq z\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\Big(\frac{-t^2}{2}\Big) dt = \Phi(z) \quad as \quad n \to \infty.$$

C) Lindeberg-Feller condition:

$$\forall \varepsilon > 0, \quad \frac{1}{s_n^2} \sum_{k=1}^{n} \mathrm{E}\Big[(X_k - \mu_k)^2 I_{\{|X_k - \mu_k| > \varepsilon s_n\}}\Big] \to 0 \quad as \quad n \to \infty.$$

Then, (A) and (B) hold simultaneously if and only if (C) holds.

Theorem 1.3.3 (Slutsky) *Let* $\{X_n\}$ *and* $\{Y_n\}$ *be sequences of random variables such that* $X_n \xrightarrow{\mathcal{D}} X$ *and* $Y_n \xrightarrow{P} c$, *where c is a constant. Then, it follows that*

i) $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$,
ii) $Y_n X_n \xrightarrow{\mathcal{D}} cX$,
iii) $X_n/Y_n \xrightarrow{\mathcal{D}} X/c \quad if \quad c \neq 0$.

Theorem 1.3.4 (Khintchine Strong Law of Large Numbers) *Let* $X_i, i \geq 1$ *be independently identically distributed random variables. Then* $\bar{X}_n \xrightarrow{a.s.} c$, *and only if* $\mathrm{E}(X_1)$ *exists and* $c = \mathrm{E}(X_1)$.

Theorem 1.3.5 (Delta Method) *Let* $\{\mathrm{T}_n\}$ *be a sequence of random vectors such that* $\sqrt{n}(\mathrm{T}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \Sigma)$ *and consider a real-valued function* $g(\mathrm{T}_n)$ *such that* $g'(\theta)$ *is non-null and continuous in a neighborhood of* θ. *Then*

$$\sqrt{n}[g(\mathrm{T}_n) - g(\theta)] \xrightarrow{\mathcal{D}} N(0, \gamma^2) \quad \text{with} \quad \gamma^2 = [g'(\theta)]^T \Sigma [g'(\theta)].$$

Chapter 2

Generalized linear models and logistic regression model with binary data

2.1 Introduction

As we intend to construct simultaneous confidence bands for both linear and logistic regression models, it is motivated to introduce the generalized linear models first. The so-called generalized linear models, an extension of the classical linear modelling process that allows models to be fitted to data, can be analogously used in the following more general situations: first, the response variables have probability distributions other than the normal distribution, such as poisson, binomial, multinomial and etc; second, the relationship between the response and the predictor variables are not necessarily of the linear form. Also, generalized linear models relax the requirement of equality or constancy of variances that is required for hypothesis testing in traditional linear models. Generalized linear models include, as special cases, the linear regression and analysis of variance models, the log-linear models for categorical data, the product multinomial response models, the logistic model with binary data as well as some simple statistical models arising in survival analysis. In particular, the logistic regression model with a binary response variable is of our interest in this thesis.

In this chapter, we first specify the models, followed by the consideration of the parameter estimation based on the maximum likelihood and Newton-Raphson iterative method. Then, we focus on the asymptotic behavior of the estimators. Some related statistical inferences are considered after that specially for the logistic regression model.

2.2 Model specification

2.2.1 Generalized linear models

Consider a single random variable Y whose probability distribution depends on a single parameter θ. The distribution belongs to the exponential family if it can be written of the form given by

$$f(y, \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)], \tag{2.1}$$

where a, b, c and d are known functions. Specially, if $a(y) = y$, the distribution is said to be in canonical form. If there are other parameters, in addition to the parameter of interest θ, they are regarded as nuisance parameters forming parts of the functions a, b, c and d, and they are treated as though they are known. Many familiar distributions belong to the exponential family. For example, the poisson distribution, the normal distribution and the binomial distribution can all be written in the canonical form. Details can be found in, e.g., Dobson (2001).

The idea of a generalized linear model was introduced by Nelder and Wedderburn (1972) to demonstrate a unity of many statistical methods. This model is defined in terms of a set of independent random variables $Y_1, \ldots, Y_N$ each with a distribution from the exponential family and has the following properties:

1. the distribution of each Y_i has the canonical form and depends on a

single parameter θ_i, thus

$$f(y_i, \theta_i) = \exp[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)]; \tag{2.2}$$

2. the distribution of all Y_i's are of the same form so that the subscripts on b, c, and d can all be ignored, thus the joint probability density function of $Y_1, \dots, Y_N$ is given by

$$\begin{aligned} & f(y_1, \dots, y_N; \theta_1, \dots, \theta_N) \\ = & \prod_{i=1}^{N} \exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)] \\ = & \exp\Big[\sum_{i=1}^{N} y_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)\Big]. \end{aligned} \tag{2.3}$$

The parameters θ_i and the observations of $y_i, i = 1, \dots, N$ may one-to-one correspond, which leads that θ_i's are typically not of direct interest. A smaller set of parameters $\beta_1, \dots, \beta_p$ (where $p < N$) are usually adopted. Suppose that $\mathrm{E}(Y_i) = \mu_i$ where μ_i is some function of θ_i. In a generalized linear model, a relationship between μ_i and a linear combination $\mathbf{x}_i^T\boldsymbol{\beta}$ is specified as

$$g(\mu_i) = \mathbf{x}_i^T\boldsymbol{\beta}, \tag{2.4}$$

where g is a monotone and differentiable function called the link function, $\mathbf{x}_i$ is a p-dimensional vector of the predictor variables and the ith column of the design matrix X as well, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is the parameter vector of interest. Different link function determines different class of generalized linear models the considered model belongs to.

2.2.2 Binary response and logistic regression model

In this subsection, we consider a generalized linear model in which the outcome variable is measured on a binary scale. 'Success' and 'failure' are usually used as generic terms of the two categories.

Define the binary random variable

$$Z = \begin{cases} 1 & \text{if the outcome is a success,} \\ 0 & \text{if the outcome is a failure,} \end{cases}$$

with probabilities $\mathrm{P}\{Z = 1\} = \pi$ and $\mathrm{P}\{Z = 0\} = 1 - \pi$. If there are n such random variables $Z_1, \ldots, Z_n$ which are independent and with $\mathrm{P}\{Z_j = 1\} = \pi_j$, then their joint probability is

$$\prod_{j=1}^{n} \pi_j^{z_j}(1-\pi_j)^{1-z_j} = \exp\Big[\sum_{j=1}^{n} z_j \log\big(\frac{\pi_j}{1-\pi_j}\big) + \sum_{j=1}^{n} \log(1-\pi_j)\Big], \qquad (2.5)$$

which obviously is a member of the exponential family. For the case when π_j's are all equal, a new random variable can be defined by

$$Y = \sum_{j=1}^{n} Z_j$$

so that Y is the number of successes in n 'trials'. Then Y has the binomial distribution with parameters n and π, and its probability distribution function is given by

$$\mathrm{P}\{Y = y\} = \begin{pmatrix} n \\ y \end{pmatrix} \pi^y (1-\pi)^{n-y}, \qquad y = 0, 1, \ldots, n.$$

Now consider m independent such random variables $Y_1, \ldots, Y_m$ corresponding to the numbers of successes in m different subgroups. Each subgroup is of size $n_i, i = 1, \ldots, m$ such that $\sum_{i=1}^{m} n_i = N$. Since $Y_i \sim$ *binomial*(n_i, π_i), the log-likelihood function is therefore given by

$$\begin{aligned} & l(\pi_1, \ldots, \pi_m; y_1, \ldots, y_m) \\ = & \sum_{i=1}^{m} \left[y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + n_i \log(1-\pi_i) + \log\begin{pmatrix} n_i \\ y_i \end{pmatrix} \right]. \end{aligned} \qquad (2.6)$$

The proportion of the successes in each subgroup, i.e., $P_i = y_i/n_i$, $i = 1, \ldots, m$, is of interest. Note that $\mathrm{E}(Y_i) = n_i\pi_i$ implies $\mathrm{E}(P_i) = \pi_i$. The probability π_i is linked with the parameters of interest by

$$g(\pi_i) = \mathrm{x}_i^T \beta,$$

where $\mathbf{x}_i$ and $\boldsymbol{\beta}$ are the same as before, g is the link function. Therefore, the general linear logistic regression model is defined by setting the link function

$$g = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T\boldsymbol{\beta}, \quad i = 1, \ldots, m, \tag{2.7}$$

where $\log[\pi_i/(1-\pi_i)]$ is sometimes called the logit function.

2.3 Parameter estimation

To estimate parameters in a generalized linear model, we use a method based on the maximum likelihood. Although explicit mathematical expression can be obtained for the estimators of the parameters in some special cases, numerical method is usually needed which is typically iterative and based on the Newton-Raphson algorithm.

Consider the independent random variables $Y_1, \ldots, Y_m$ that fulfil the requirements of a generalized linear model. We have $\mathrm{E}(Y_i) = \mu_i$ and $g(\mu_i) = \mathbf{x}_i^T\boldsymbol{\beta} = \eta_i$, where $\mathbf{x}_i$ is the vector with the elements x_{ij}, $i = 1, \ldots, m$ indicating which subgroup the observation belongs to and $j = 1, \ldots, p$ indicating which predictor variable is observed. For each Y_i, the likelihood function is

$$f(y_i, \theta_i) = \exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)], \tag{2.8}$$

where the functions b, c and d are known. In order to derive the score functions as well as the information matrix, expressions for the expected value and variance of Y_i's are needed. The following method is used to find the score functions and the information matrix by changing the order of integration and differentiation provided a density function.

By the property that a probability density function integrates to 1, we have

$$\frac{\partial}{\partial\theta_i}\int f(y_i, \theta_i)\mathrm{d}y_i = \frac{\partial}{\partial\theta_i}\cdot 1 = 0. \tag{2.9}$$

Changing the order of the integration and differentiation, (2.9) becomes

$$\int \frac{\partial f(y_i, \theta_i)}{\partial\theta_i}\mathrm{d}y_i = 0. \tag{2.10}$$

Similarly, when the differentiation in (2.9) is of second order, then we have

$$\int \frac{\partial^2 f(y_i, \theta_i)}{\partial \theta_i^2} \mathrm{d}y_i = 0. \tag{2.11}$$

These results can be used to obtain the expectations and the variances of Y_i's. (2.10) can be further written as

$$\begin{aligned} & \int f(y_i, \theta_i)[y_i b'(\theta_i) + c'(\theta_i)] \\ = & \int f(y_i, \theta_i) y_i b'(\theta_i) + \int f(y_i, \theta_i) c'(\theta_i) \\ = & \ b'(\theta_i)\mu_i + c'(\theta_i) = 0 \end{aligned}$$

Thus, we have

$$\mathrm{E}(Y_i) = \mu_i = -c'(\theta_i)/b'(\theta_i). \tag{2.12}$$

Similarly, we have

$$\mathrm{Var}(Y_i) = [b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)]/[b'(\theta_i)]^3. \tag{2.13}$$

Now we turn to derive the score function and the information matrix. The log-likelihood function for all the Y_i's is

$$l = \sum_{i=1}^{m} l_i = \sum_{i=1}^{m} y_i b(\theta_i) + \sum_{i=1}^{m} c(\theta_i) + \sum_{i=1}^{m} d(y_i). \tag{2.14}$$

We use the chain rule for differentiation to obtain the score function which is given by

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{m} \Big(\frac{\partial l_i}{\partial \beta_j}\Big) = \sum_{i=1}^{m} \Big(\frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \beta_j}\Big). \tag{2.15}$$

Consider each term on the right-hand side of (2.15) separately. $\partial l_i/\partial \theta_i$ can be obtained from (2.14), $\partial \theta_i/\partial \mu_i$ can be obtained from (2.12), and $\partial \mu_i/\partial \beta_j$ can be obtained from the link function. Substituting these three individuals into (2.15) finally gives

$$U_j = \sum_{i=1}^{m} \Big[\frac{(y_i - \mu_i)}{\mathrm{Var}(Y_i)} x_{ij} \Big(\frac{\partial \mu_i}{\partial \eta_i}\Big)\Big]. \tag{2.16}$$

The variance-covariance matrix of U_j has the terms

$$\begin{aligned} J_{jk} &= \mathrm{E}(U_j U_k) \\ &= \sum_{i=1}^{m} \Big[\frac{x_{ij}x_{ik}}{\mathrm{Var}(Y_i)}\Big(\frac{\partial \mu_i}{\partial \eta_i}\Big)^2\Big], \end{aligned} \tag{2.17}$$

where $j, k = 1, \ldots, p$.

An iterative procedure is usually adopted for parameter estimation. The most commonly used method is the following Newton-Raphson approximation. Suppose t is a function of x, Newton-Raphson method is the iterative process described by

$$x^{(r)} = x^{(r-1)} - \frac{t\Big(x^{(r-1)}\Big)}{t'\Big(x^{(r-1)}\Big)} \tag{2.18}$$

to find the value of x such that $t(x) = 0$. It starts with an initial guess $x^{(1)}$ to obtain successive approximation until the iterative process converges.

By Newton-Raphson's formula, the rth approximation of the parameter vector β is given by

$$\mathrm{b}^{(r)} = \mathrm{b}^{(r-1)} - \Big(\frac{\partial^2 l}{\partial\beta\partial\beta^T}\Big)^{-1}_{\beta=\mathrm{b}^{(r-1)}} \mathrm{U}^{(r-1)}, \tag{2.19}$$

where $\mathrm{b}^{(r)}$ denotes the vector of the estimates of the parameter vector β at the rth iteration, $\mathrm{U}^{(r-1)}$ is the vector of the first order derivatives U_j's evaluated at $\beta = \mathrm{b}^{(r-1)}$. By the method of scoring which replaces the matrix of the second order derivatives in (2.19) by its expectation, and the fact that

$$J = -\mathrm{E}\Big(\frac{\partial^2 l}{\partial\beta\partial\beta^T}\Big),$$

we have (2.19) equal to

$$\mathrm{b}^{(r)} = \mathrm{b}^{(r-1)} + \Big[\mathcal{J}^{(r-1)}\Big]^{-1}\mathrm{U}^{(r-1)}, \tag{2.20}$$

where $[\mathcal{J}^{(r-1)}]^{-1}$ is the inverse of the information matrix with the elements J_{jk} given by (2.17) all evaluated at $\mathrm{b}^{(r-1)}$. An alternative version gives

$$\mathcal{J}^{(r-1)}\mathrm{b}^{(r)} = \mathcal{J}^{(r-1)}\mathrm{b}^{(r-1)} + \mathrm{U}^{(r-1)}. \tag{2.21}$$

By (2.17) J can be written as

$$J = X^T WX,$$

where W is an $m \times m$ diagonal matrix with the elements given by

$$w_{ii} = \frac{1}{\text{Var}(Y_i)}\Big(\frac{\partial \mu_i}{\partial \eta_i}\Big)^2. \tag{2.22}$$

By (2.16) and (2.17), the right-hand side of the equation (2.21) can be written as $X^T W\mathbf{z}$, where $\mathbf{z}$ has the elements

$$z_i = \sum_{j=1}^{p} x_{ij} b_j^{(r-1)} + (y_i - \mu_i)\Big(\frac{\partial \eta_i}{\partial \mu_i}\Big) \tag{2.23}$$

with μ_i and $\partial\eta_i/\partial\mu_i$ evaluated at $\mathrm{b}^{(r-1)}$. Hence the iterative equation (2.21) is equal to

$$X^T W^{(r-1)} X\mathrm{b}^{(r)} = X^T W^{(r-1)}\mathbf{z}^{(r-1)} \tag{2.24}$$

which has to be solved iteratively because, in general, $\mathbf{z}$ and W depend on b. Thus for generalized linear models, the maximum likelihood estimates are obtained by an iterative weighted procedure.

In particular for logistic regression model, we have the log-likelihood function given by

$$l(\boldsymbol{\pi}; \mathbf{y}) = \sum_{i=1}^{m}\Big[y_i \log\Big(\frac{\pi_i}{1-\pi_i}\Big) + n_i \log(1-\pi_i) + \log\binom{n_i}{y_i}\Big],$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)^T$ and $\mathbf{y} = (y_1, \ldots, y_m)^T$. Also, we have the link function

$$g(\pi_i) = \eta_i = \log\Big(\frac{\pi_i}{1-\pi_i}\Big) = \mathrm{x}_i^T \boldsymbol{\beta}.$$

Using the chain rule in (2.15), we have

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{m}\Big(\frac{\partial l_i}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j}\Big) = \sum_{i=1}^{m}(y_i - n_i\pi_i)x_{ij}.$$

The fisher information for $\boldsymbol{\beta}$ is therefore

$$J_{jk} = -\mathrm{E}\Big(\frac{\partial^2 l}{\partial\beta_j \partial\beta_k}\Big) = \sum_{i=1}^{m} n_i\pi_i(1-\pi_i)x_{ij}x_{ik} = \{X^T WX\}_{jk},$$

where $j, k = 1, \ldots, p$ and W is a diagonal matrix of the weights given by

$$W = \text{diag}\{n_i \pi_i (1 - \pi_i)\}.$$

Following the Newton-Raphson procedure, define $\mathbf{z}$ with the elements given by

$$z_i = \hat{\eta}_i + \frac{y_i - n_i \hat{\pi}_i}{n_i} \cdot \frac{\partial \eta_i}{\partial \pi_i},$$

then the maximum likelihood estimates can be obtained from the equation (2.24).

Most statistical packages include the algorithm of estimation for generalized linear models. They begin by evaluating $\mathbf{z}$ and W using some initial approximation $\mathbf{b}^{(0)}$, then solve the iterative equation (2.24) to obtain $\mathbf{b}^{(1)}$ which in turn is used to get better approximations for $\mathbf{z}$ and W, and so on until adequate convergence is reached. When the difference between the two successive approximations $\mathbf{b}^{(r-1)}$ and $\mathbf{b}^{(r)}$ is sufficiently small, then $\mathbf{b}^{(r)}$ is taken as the maximum likelihood estimate of the parameter vector $\boldsymbol{\beta}$.

2.4 Asymptotic behavior of estimators

2.4.1 Introduction

Since most distributional inferences on generalized linear models are valid based on large samples, there is a need to look into the large sample asymptotic theory so that some desired distributional properties for the estimators can be obtained. Specifically, the asymptotic normality of the maximum likelihood estimators is of interest.

Recall the specification of a generalized linear model in Section 2.2. Consider the vector of observations $\mathbf{y} = (y_1, \ldots, y_m)^T$ corresponding to m independent random variables $Y_i, i = 1, \ldots, m$, each with a distribution in the exponential family. We have the density function for each Y_i given by

$$f(y_i, \theta_i, \phi) = c(y_i, \phi) \exp\{[y_i \theta_i - b(\theta_i)]/a(\phi)\}, \tag{2.25}$$

where θ_i's are parameters, $\phi > 0$ is a scale and $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are all known functions which are distinguishable from those appeared previously. Therefore the joint density function is

$$\begin{aligned} & f(y_1, \ldots, y_m; \theta_1, \ldots, \theta_m; \phi) \\ = & \prod_{i=1}^{m} f(y_i, \theta_i, \phi) \\ = & c(\mathbf{y}, \phi)\exp\Big\{ \sum_{i=1}^{m} [y_i\theta_i - b(\theta_i)]/a(\phi) \Big\}, \end{aligned} \tag{2.26}$$

where

$$c(\mathbf{y}, \phi) = \prod_{i=1}^{m} c(y_i, \phi).$$

Review that

$$\mathrm{E}(Y_i) = \mu_i(\theta_i) = b'(\theta_i), \tag{2.27}$$

$$\mathrm{Var}(Y_i) = a(\phi)b''(\theta_i) = a(\phi)v_i[\mu_i(\theta_i)], \tag{2.28}$$

where $v_i[\mu_i(\theta_i)]$ is known as the variance function of θ_i which depends solely on $\mu_i(\theta_i)$ for $1 \leq i \leq m$. Furthermore, conceive of a transformation which provides the link between μ_i and $\mathbf{x}_i^T\boldsymbol{\beta}$ of the form

$$g[\mu(\theta_i)] = \mathbf{x}_i^T\boldsymbol{\beta}, \quad i = 1, \ldots, m, \tag{2.29}$$

where $g(\cdot)$ is a monotone and differentiable function and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the p-dimensional parameter vector. Alternatively, the link can be arranged in multi-dimensional version that

$$\mathbf{G} = \Big(g(\mu_1), \ldots, g(\mu_m)\Big)^T = X\boldsymbol{\beta}, \tag{2.30}$$

where $X = (\mathbf{x}_1, \ldots, \mathbf{x}_m)^T$ denotes a known $m \times p$ matrix.

Now we turn to find the asymptotic distribution of $\hat{\boldsymbol{\beta}}$, the maximum likelihood estimator of $\boldsymbol{\beta}$ in the generalized linear model. Usually it is assumed in the asymptotic sense that the total sample size $N \to \infty$, where $N = \sum_{i=1}^{m} n_i$ with n_i being the sample size of the ith subgroup of observations. However, there may be another situation where for each i, the Y_i may be a statistic given the subsample size n_i. In such a case, a second type of asymptotic might be considered where it is not crucial to have N large, provided the n_i's are themselves large. In the rest of this section, we focus on the regular case first, where $N \to \infty$, and briefly talk about the second type of asymptotic after that.

2.4.2 The first type of asymptotic

Define $h(\cdot) = (g \circ \mu)^{-1}(\cdot)$ so that (2.29) is transformed to

$$\theta_i = h(\mathbf{x}_i^T \beta), \quad i = 1, \ldots, m, \tag{2.31}$$

where h is monotone and differentiable. The parameter vector β is of direct interest. By reviewing (2.25) and (2.26), we may note that the nuisance parameter ϕ does not affect the estimation of β and it influences the information matrix J only by a multiplicative factor $[a(\phi)]^{-2}$ which may be estimated consistently. Therefore, for the sake of simplicity and without loss of generality, $a(\phi) \equiv 1$ is taken. Consider (2.31), the log-likelihood function in terms of β is given by

$$\log L_N(\beta) = \sum_{i=1}^{m} \left\{ n_i y_i h(\mathbf{x}_i^T \beta) - n_i b[h(\mathbf{x}_i^T \beta)] \right\} - \text{constant}, \tag{2.32}$$

where the constant term does not depend on β, the subscript of the likelihood, i.e., N, indicates that it is for the first type of asymptotic, and the quantities with such a subscript hereafter in this chapter are of the same meaning. Recall (2.27) and (2.28), then we have

$$\mu_i(\beta) = \mu[h(\mathbf{x}_i^T \beta)] = b'[h(\mathbf{x}_i^T \beta)], \tag{2.33}$$

$$v_i(\beta) = v[h(\mathbf{x}_i^T \beta)] = b''[h(\mathbf{x}_i^T \beta)]. \tag{2.34}$$

Thus, the score function of β is given by

$$\mathbf{U}_N(\beta) = \frac{\partial}{\partial \beta} \log L_N(\beta) = \sum_{i=1}^{m} \frac{n_i[y_i - \mu_i(\beta)]}{\{g'[\mu_i(\beta)] v_i(\beta)\}} \mathbf{x}_i, \tag{2.35}$$

from which it follows that whenever $g'(\cdot)$ and $v(\cdot) \equiv b''(\cdot)$ are both differentiable, then we have

$$-\frac{\partial^2}{\partial\beta\partial\beta^T}\log L_N(\beta) = J_N(\beta) + R_N(\beta), \tag{2.36}$$

where

$$J_N(\beta) = \sum_{i=1}^{m} n_i\{g'[\mu_i(\beta)]\}^{-2}[v_i(\beta)]^{-1}\mathbf{x}_i\mathbf{x}_i^T \tag{2.37}$$

and

$$\begin{aligned} R_N(\beta) &= \sum_{i=1}^{m} n_i[y_i - \mu_i(\beta)] \\ &\quad \times\Big\{\frac{g''[\mu_i(\beta)]}{\{g'[\mu_i(\beta)]\}^2} + \frac{b'''[h(\mathbf{x}_i^T\beta)]}{\{g'[\mu_i(\beta)]\}^2[v_i(\beta)]^3}\Big\}\mathbf{x}_i\mathbf{x}_i^T. \end{aligned} \tag{2.38}$$

Remark 2.4.1 *Recall (2.31) that when $g = \mu^{-1}$, $g \circ \mu$ is the identity function, hence, $\theta_i = \mathbf{x}_i^T\beta$. In such a case, $g(\cdot)$ is termed a canonical link function. By (2.36)-(2.38), we have that for canonical link functions, $R_N(\beta) = 0$.*

Proof. For canonical link functions, we have $h(\mathbf{x}_i^T\beta) = \mathbf{x}_i^T\beta$ and $g = \mu^{-1}$. Thus

$$\begin{aligned} & b'[h(\mathbf{x}_i^T\beta)] = \mu(\mathbf{x}_i^T\beta) = g^{-1}(\mathbf{x}_i^T\beta) \\ \Rightarrow\ & g\{b'[h(\mathbf{x}_i^T\beta)]\} = \mathbf{x}_i^T\beta \\ \Rightarrow\ & g[b'(\mathbf{x}_i^T\beta)] = \mathbf{x}_i^T\beta. \end{aligned} \tag{2.39}$$

Differentiate the both sides of the last equality in (2.39) with respect to β and obtain

$$g'[b'(\mathbf{x}_i^T\beta)] \cdot b''(\mathbf{x}_i^T\beta) \cdot \mathbf{x}_i = \mathbf{x}_i \tag{2.40}$$

which implies

$$b''(\mathbf{x}_i^T\beta) = \{g'[b'(\mathbf{x}_i^T\beta)]\}^{-1}. \tag{2.41}$$

Differentiate twice and obtain

$$g''[b'(\mathbf{x}_i^T\beta)] \cdot [b''(\mathbf{x}_i^T\beta)]^2 \cdot \mathbf{x}_i\mathbf{x}_i^T + g'[b'(\mathbf{x}_i^T\beta)] \cdot b'''(\mathbf{x}_i^T\beta) \cdot \mathbf{x}_i\mathbf{x}_i^T = 0 \tag{2.42}$$

which, in connection with (2.41), implies

$$\begin{aligned}
&\Big\{\frac{g''[b'(\mathbf{x}_i^T\boldsymbol{\beta})]}{\{g'[b'(\mathbf{x}_i^T\boldsymbol{\beta})]\}^2}+\frac{b'''(\mathbf{x}_i^T\boldsymbol{\beta})}{g'[\mu_i(\boldsymbol{\beta})]\cdot[b''(\mathbf{x}_i^T\boldsymbol{\beta})]^2}\Big\}\mathbf{x}_i\mathbf{x}_i^T\\
=\ &\Big\{\frac{g''[b'(\mathbf{x}_i^T\boldsymbol{\beta})]}{\{g'[b'(\mathbf{x}_i^T\boldsymbol{\beta})]\}^2}+\frac{b'''(\mathbf{x}_i^T\boldsymbol{\beta})}{\{g'[\mu_i(\boldsymbol{\beta})]\}^2[b''(\mathbf{x}_i^T\boldsymbol{\beta})]^3}\Big\}\mathbf{x}_i\mathbf{x}_i^T=0. \qquad (2.43)
\end{aligned}$$

We therefore simply obtain $R_N(\boldsymbol{\beta})=0$. #

In order to obtain the asymptotic distribution of the estimator of $\boldsymbol{\beta}$ under the consistent setup, we need to discuss some required assumptions.

Assumption 2.4.1 *Assume that*

$$\lim_{N\to\infty}\frac{1}{N}J_N(\boldsymbol{\beta})=J(\boldsymbol{\beta}),\ \textit{finite and positive definite.} \qquad (2.44)$$

Assumption 2.4.2 *Let*

$$\mathbf{G}_i=\Big\{\frac{g''[\mu_i(\boldsymbol{\beta})]}{(g'[\mu_i(\boldsymbol{\beta})])^2}+\frac{b'''[h(\mathbf{x}_i^T\boldsymbol{\beta})]}{\{g'[\mu_i(\boldsymbol{\beta})]\}^2[v_i(\boldsymbol{\beta})]^3}\Big\}\mathbf{x}_i\mathbf{x}_i^T=w_i\mathbf{x}_i\mathbf{x}_i^T, \qquad (2.45)$$

where

$$w_i=\Big\{\frac{g''[\mu_i(\boldsymbol{\beta})]}{\{g'[\mu_i(\boldsymbol{\beta})]\}^2}+\frac{b'''[h(\mathbf{x}_i^T\boldsymbol{\beta})]}{\{g'[\mu_i(\boldsymbol{\beta})]\}^2[v_i(\boldsymbol{\beta})]^3}\Big\},\quad i=1,\ldots,m, \qquad (2.46)$$

and assume that

$$\lim_{N\to\infty}N^{-2}\sum_{i=1}^{m}n_i^2v_i[\mu_i(\boldsymbol{\beta})]\mathrm{tr}(\mathbf{G}_i\mathbf{G}_i^T)=0. \qquad (2.47)$$

By directly applying Chebyshev Inequality (1.11) to (2.38) that for $\forall\varepsilon>0$,

$$\begin{aligned}
&\mathrm{P}\{N^{-1}R_N(\boldsymbol{\beta})>\varepsilon\}\\
=\ &\mathrm{P}\{[N^{-1}R_N(\boldsymbol{\beta})]^2>\varepsilon^2\}\\
\le\ &\frac{\mathrm{E}[N^{-1}R_N(\boldsymbol{\beta})]^2}{\varepsilon^2}\\
=\ &\frac{N^{-2}\sum_{i=1}^m n_i^2v_i[\mu_i(\boldsymbol{\beta})]\mathrm{tr}(\mathbf{G}_i\mathbf{G}_i^T)}{\varepsilon^2}, \qquad (2.48)
\end{aligned}$$

in connection with Assumption 2.4.2, we can simply obtain $N^{-1}R_N(\beta) \xrightarrow{P} 0$. From (2.36), we have

$$\frac{1}{N}\Big[\frac{\partial^2}{\partial\beta\partial\beta^T}\log L_N(\beta) + J_N(\beta)\Big] \xrightarrow{P} 0. \tag{2.49}$$

Next, we find the asymptotic normality of the score function given in (2.35). Observe that $\mathrm{E}[\mathbf{U}_N(\beta)] = 0$ and $\mathrm{Var}[\mathbf{U}_N(\beta)] = J_N(\beta)$. To apply Lindeberg-Feller Theorem to the independent but not necessarily identically distributed random vectors, it is needed to show the Lindeberg-Feller condition is satisfied. For $\forall\ \varepsilon > 0$, consider

$$J_N^{-1}(\beta)\sum_{i=1}^{m}\mathrm{E}(\mathbf{t}_i\mathbf{t}_i^T)\mathrm{P}\{\|\mathbf{t}_i\mathbf{t}_i^T\| > \varepsilon\|J_N(\beta)\|\}, \tag{2.50}$$

where $\mathbf{t}_i$ is defined such that

$$\mathbf{U}_N(\beta) = \sum_{i=1}^{m}\frac{n_i[y_i - \mu_i(\beta)]}{\{g'[\mu_i(\beta)]v_i(\beta)\}}\mathbf{x}_i = \sum_{i=1}^{m}\mathbf{t}_i. \tag{2.51}$$

By applying Chebyshev Inequality, we have

$$\begin{aligned}
& J_N^{-1}(\beta)\sum_{i=1}^{m}\mathrm{E}(\mathbf{t}_i\mathbf{t}_i^T)\mathrm{P}\{\|\mathbf{t}_i\mathbf{t}_i^T\| > \varepsilon\|J_N(\beta)\|\} \\
\leq\ & J_N^{-1}(\beta)\sum_{i=1}^{m}\mathrm{E}(\mathbf{t}_i\mathbf{t}_i^T)\frac{\mathrm{E}\|\mathbf{t}_i\mathbf{t}_i^T\|}{\varepsilon\|J_N(\beta)\|} \qquad (2.52) \\
\leq\ & \frac{\max(\mathrm{E}\|\mathbf{t}_i\mathbf{t}_i^T\|)}{\varepsilon\|J_N(\beta)\|}J_N^{-1}(\beta)\sum_{i=1}^{m}\mathrm{E}(\mathbf{t}_i\mathbf{t}_i^T), \qquad (2.53)
\end{aligned}$$

where $\max(\mathrm{E}\|\mathbf{t}_i\mathbf{t}_i^T\|)$ is the maximum of $\mathrm{E}\|\mathbf{t}_i\mathbf{t}_i^T\|$ as $i = 1, 2, \ldots, m$. Also it is known that $J_N^{-1}(\beta)\sum_{i=1}^{m}\mathrm{E}(\mathbf{t}_i\mathbf{t}_i^T)$ is equal to the identity matrix. From Assumption 2.4.1, we know that $\|J_N(\beta)\| \to \infty$ as $N \to \infty$. Thus, we have that the right-hand side of (2.53) converges to zero as $N \to \infty$ so that (2.50) converges to zero as $N \to \infty$. Lindeberg-Feller condition is satisfied. By Lindeberg-Feller Theorem in connection with Assumption 2.4.1, we therefore have the asymptotic normality of $\mathbf{U}_N(\beta)$ that

$$\frac{1}{\sqrt{N}}\mathbf{U}_N(\beta) \xrightarrow{\mathcal{D}} N(0, J(\beta)). \tag{2.54}$$

Now we turn to our central work to show the asymptotic normality of the maximum likelihood estimator $\hat{\beta}$. Let $\|\mathrm{u}\| < k, 0 < k < \infty$ and consider the Taylor Expansion of $\log L_N(\beta + N^{-\frac{1}{2}}\mathrm{u})$ around $\log L_N(\beta)$. Define

$$\begin{aligned}\lambda_N(\mathrm{u}) &= \log L_N(\beta + N^{-\frac{1}{2}}\mathrm{u}) - \log L_N(\beta) \\ &= \frac{1}{\sqrt{N}}\mathrm{u}^T \mathbf{U}_N(\beta) + \frac{1}{\sqrt{2N}}\mathrm{u}^T \frac{\partial^2}{\partial\beta\partial\beta^T}\log L_N(\beta)\Big|_{\beta^*}\mathrm{u}, \quad (2.55)\end{aligned}$$

where β^* is a point belonging to the line ended by β and $(\beta + N^{-\frac{1}{2}}\mathrm{u})$ in the parameter space. Also, define

$$\begin{aligned}Z_N(\mathrm{u}) &= \frac{1}{\sqrt{2N}}\Big\{\mathrm{u}^T \frac{\partial^2}{\partial\beta\partial\beta^T}\log L_N(\beta)\Big|_{\beta^*}\mathrm{u} - \mathrm{u}^T \frac{\partial^2}{\partial\beta\partial\beta^T}\log L_N(\beta)\Big|_{\beta}\mathrm{u} \\ &\quad + \mathrm{u}^T \frac{\partial^2}{\partial\beta\partial\beta^T}\log L_N(\beta)\Big|_{\beta}\mathrm{u} + \mathrm{u}^T J_N(\beta)\mathrm{u}\Big\}. \quad (2.56)\end{aligned}$$

Then (2.55) can be alternatively written as

$$\lambda_N(\mathrm{u}) = \frac{1}{\sqrt{N}}\mathrm{u}^T \mathbf{U}_N(\beta) - \frac{1}{\sqrt{2N}}\mathrm{u}^T J_N(\beta)\mathrm{u} + Z_N(\mathrm{u}). \quad (2.57)$$

Observe that

$$\begin{aligned}&\sup_{\|\mathrm{u}\|<k} \|Z_N(\mathrm{u})\| \\ \leq\ & \frac{1}{2}\sup_{\beta^* \in \mathrm{B}(k/\sqrt{N})}\Big\|\frac{1}{N}\frac{\partial^2}{\partial\beta\partial\beta^T}\log L_N(\beta)\Big|_{\beta^*} - \frac{1}{N}\frac{\partial^2}{\partial\beta\partial\beta^T}\log L_N(\beta)\Big|_{\beta}\Big\| \\ & + \frac{k^2}{2}\Big\|\frac{1}{N}\frac{\partial^2}{\partial\beta\partial\beta^T}\log L_N(\beta)\Big|_{\beta} + \frac{1}{N}J_N(\beta)\Big\|, \quad (2.58)\end{aligned}$$

where $\mathrm{B}(k/\sqrt{N})$ is defined in the following assumption. From (2.49), the second absolute value on the right-hand side of (2.58) converges to zero in probability as $N \to \infty$. Next we consider the first term on the right-hand side of (2.58).

Assumption 2.4.3 *Suppose that* $\{g'[\mu_i(\beta^*)]\}^{-2}$, $[v_i(\beta^*)]^{-3}$, $g''[\mu_i(\beta^*)]$ *and* $b'''[h(\mathrm{x}_i^T\beta^*)]$ *are uniformly continuous in an infinitesimal neighborhood of the true* β, *i.e., in the set*

$$\mathrm{B}(\delta) = \{\beta^* \in \mathcal{R}^p : \|\beta^* - \beta\| < \delta\}, \quad \delta \downarrow 0. \quad (2.59)$$

Let

$$w_{1i}(\beta) = \{g'[\mu_i(\beta)]\}^{-2}[v_i(\beta)]^{-1}$$

and

$$w_{2i}(\beta) = \mu_i(\beta)w_i(\beta),$$

where $w_i(\beta)$ *is defined in (2.46),* $i = 1, \ldots, m$.

1. *For* $k = 1, 2$, *as* $\delta \downarrow 0$,

$$\sup_{\beta^* \in \mathrm{B}(\delta)} \|\{w_{ki}(\beta^*) - w_{ki}(\beta)\}\mathbf{x}_i\mathbf{x}_i^T\| \to 0. \tag{2.60}$$

2. *As* $\delta \downarrow 0$,

$$\mathrm{E}_\beta\{\sup_{\beta^* \in \mathrm{B}(\delta)} |y_i| \|w_i(\beta)^* - w_i(\beta)\mathbf{x}_i\mathbf{x}_i^T\|\} = \psi \to 0, \tag{2.61}$$

where ψ *is a scale.*

Observe according to (2.36)-(2.38) that

$$\begin{aligned}
& \mathrm{E}_\beta \left\{ \sup_{\beta^* \in \mathrm{B}(k/\sqrt{N})} \left\| \frac{\partial^2}{\partial\beta\partial\beta^T} \log f(y_i;\beta)\Big|_{\beta^*} - \frac{\partial^2}{\partial\beta\partial\beta^T} \log f(y_i;\beta)\Big|_{\beta} \right\| \right\} \\
= \; & \mathrm{E}_\beta \left\{ \sup_{\beta^* \in \mathrm{B}(k/\sqrt{N})} \left\| \{w_{1i} + [y_i - \mu_i(\beta)]w_i\}\mathbf{x}_i\mathbf{x}_i^T\Big|_{\beta^*} \right. \right. \\
& \left. \left. - \{w_{1i} + [y_i - \mu_i(\beta)]w_i\}\mathbf{x}_i\mathbf{x}_i^T\Big|_{\beta} \right\| \right\} \\
\leq \; & \sup_{\beta^* \in \mathrm{B}(k/\sqrt{N})} \|[w_{1i}(\beta^*) - w_{1i}(\beta)]\mathbf{x}_i\mathbf{x}_i^T\| && (2.62) \\
& + \sup_{\beta^* \in \mathrm{B}(k/\sqrt{N})} \|[w_{2i}(\beta^*) - w_{2i}(\beta)]\mathbf{x}_i\mathbf{x}_i^T\| && (2.63) \\
& + \mathrm{E}_\beta \{ \sup_{\beta^* \in \mathrm{B}(k/\sqrt{N})} |y_i| \; \|[w_i(\beta^*) - w_i(\beta)]\mathbf{x}_i\mathbf{x}_i^T\|\}. && (2.64)
\end{aligned}$$

By Assumption 2.4.3, (2.62)-(2.64) converge to zero separately as $k/\sqrt{N} \to 0$, which implies that the left-hand side of the above inequality converges to zero as $k/\sqrt{N} \to 0$. In addition, observe that

$$\begin{aligned}
& \frac{1}{2} \sup_{\beta^* \in \mathrm{B}(k/\sqrt{N})} \left\| \frac{1}{N} \frac{\partial^2}{\partial\beta\partial\beta^T} \log L_N(\beta)\Big|_{\beta^*} - \frac{1}{N} \frac{\partial^2}{\partial\beta\partial\beta^T} \log L_N(\beta)\Big|_{\beta} \right\| \\
\leq \; & \frac{1}{2N} \sum_{i=1}^m n_i \sup_{\beta^* \in \mathrm{B}(k/\sqrt{N})} \left\| \frac{\partial^2}{\partial\beta\partial\beta^T} \log f(y_i;\beta)\Big|_{\beta^*} - \frac{\partial^2}{\partial\beta\partial\beta^T} \log f(y_i;\beta)\Big|_{\beta} \right\|.
\end{aligned} \tag{2.65}$$

By the Khintehine Strong Law of Large Numbers, since

$$\mathrm{E}_{\beta}\left\{\sup_{\beta^{*}\in\mathrm{B}(k/\sqrt{N})}\left\|\frac{\partial^{2}}{\partial\beta\partial\beta^{T}}\log f(y_{i};\beta)\Big|_{\beta^{*}}-\frac{\partial^{2}}{\partial\beta\partial\beta^{T}}\log f(y_{i};\beta)\Big|_{\beta}\right\|\right\}\to 0$$

as $k/\sqrt{N}\to 0$, we have the right-hand side of the inequality (2.65) converges to zero almost surely. And this implies the left-hand side of (2.65), which is the same as the first term on the right-hand side of (2.58), converges to zero almost surely. Consequently, from (2.58), we have

$$\sup_{\|\mathbf{u}\|<k}\|Z_{N}(\mathbf{u})\|\xrightarrow{a.s.}0. \tag{2.66}$$

Rewrite (2.57) as

$$\lambda_{N}(\mathbf{u})=\frac{1}{\sqrt{N}}\mathbf{u}^{T}\mathbf{U}_{N}(\beta)-\frac{1}{2N}\mathbf{u}^{T}J_{N}(\beta)\mathbf{u}+o_{p}(1). \tag{2.67}$$

By Assumption 2.4.1, (2.67) is equal to

$$\lambda_{N}(\mathbf{u})=\frac{1}{\sqrt{N}}\mathbf{u}^{T}\mathbf{U}_{N}(\beta)-\frac{1}{2}\mathbf{u}^{T}J(\beta)\mathbf{u}+o_{p}(1),$$

when $N\to\infty$. Maximize $\lambda_N(\mathbf{u})$ by solving $\partial\lambda_N(\mathbf{u})/\partial\mathbf{u}=0$, and then obtain

$$\hat{\mathbf{u}}=\frac{1}{\sqrt{N}}J^{-1}(\beta)\mathbf{U}_{N}(\beta)+o_{p}(1).$$

By the definition of $\lambda_N(\mathbf{u})$ in (2.55), it is clear that $\hat{\mathbf{u}}$ makes $L_N(\beta+N^{-1/2}\mathbf{u})$ maximal, and the maximum of which is obtained at $\hat{\beta}$. Therefore, we have

$$\begin{aligned}\hat{\beta} &= \beta+N^{-1/2}\hat{\mathbf{u}}+o_{p}(N^{-1/2})\\ &= \beta+N^{-1}J^{-1}(\beta)\mathbf{U}_{N}(\beta)+o_{p}(N^{-1/2}),\end{aligned} \tag{2.68}$$

which implies

$$\sqrt{N}(\hat{\beta}-\beta)=\frac{1}{\sqrt{N}}J^{-1}(\beta)\mathbf{U}_{N}(\beta)+o_{p}(1).$$

Apply Slutsky Theorem in connection with (2.54) and then obtain

$$\sqrt{N}(\hat{\beta}-\beta)\xrightarrow{\mathcal{D}}N_{p}(0,J^{-1}(\beta)). \tag{2.69}$$

The idea of how to obtain (2.69) for the first type of asymptotic comes from Sen and Singer (1993).

In particular, consider a logistic regression model. For the sake of simplicity, we assume that there is only one predictor variable and no intercept term is included in the model, i.e., $p = 1$. To avoid a degenerate binomial distribution, take $x_i \neq 0$. The canonical link function is given by

$$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right).$$

Accordingly, we have

$$g'(\pi) = \frac{1}{\pi(1-\pi)} \quad \text{and} \quad v_i(\beta) = \frac{\exp(2\beta x_i)}{[1+\exp(\beta x_i)]^2}x_i^2.$$

By Assumption 2.4.3, we have

$$\begin{aligned} w_{1i}(\beta) &= \left\{\frac{[1+\exp(\beta x_i)]^2}{\exp(\beta x_i)}\right\}^{-2} \cdot \left\{\frac{\exp(2\beta x_i)}{[1+\exp(\beta x_i)]^2}x_i^2\right\}^{-1} \\ &= \{[1+\exp(\beta x_i)]^2 x_i^2\}^{-1}. \end{aligned}$$

Considering Remark 2.4.1, the assumptions required for obtaining the asymptotic normality of the maximum likelihood estimator $\hat{\beta}$ reduce to

1. Assume that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{m} n_i[1+\exp(\beta x_i)]^{-2} = J(\beta) < \infty.$$

2. For some k: $0 < k < \infty$, as $N \to \infty$

$$\begin{aligned} &\sup_{|h|\leq k/\sqrt{N}} \left|\{1+\exp[(\beta+h)x_i]\}^{-2} - [1+\exp(\beta x_i)]^{-2}\right| \\ = &\left|\{1+\exp[(\beta+k/\sqrt{N})x_i]\}^{-2} - [1+\exp(\beta x_i)]^{-2}\right| \to 0. \end{aligned}$$

If we suppose the only predictor variable is bounded, then both conditions hold. Thus, we have

$$\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{\mathcal{D}} N(0, J^{-1}(\beta)).$$

In this case, the design matrix X reduces to the m-dimensional vector $\mathbf{x}$ such that

$$J(\beta) = \mathbf{x}^T W\mathbf{x},$$

where W is the diagonal weight matrix with the elements given by

$$w_{ii} = \frac{\exp(\beta x_i)}{[1+\exp(\beta x_i)]^2} \quad i = 1, \ldots, m,$$

x_i is the i-th element of $\mathbf{x}$.

2.4.3 The second type of asymptotic

We considered the first type of asymptotic behavior of β in last subsection, i.e., set the total sample size N large. As for the second type of asymptotic, we do not necessarily set N large and may consider the subsample sizes $n_i, i = 1, \ldots, m$ are themselves large.

Take the case of $m = 2$ for example. Consider the independent binary variables z_{ij}, $i = 1, 2, j = 1, \ldots, n_i$, which have the Bernoulli distribution and are defined by

$$z_{ij} = \begin{cases} 1 & \text{with probability} \quad \pi_i, \\ 0 & \text{with probability } 1 - \pi_i. \end{cases}$$

Then, we have $\mathrm{E}(z_{ij}) = \pi_i$ and $\mathrm{Var}(z_{ij}) = \pi_i(1-\pi_i)$. Define a random vector $\mathbf{Y}$ such that

$$Y_i = \sum_{j=1}^{n_i} \frac{z_{ij}}{n_i} \qquad i = 1, 2. \tag{2.70}$$

Then $\mathbf{Y}$ can be viewed as the vector of the frequencies of the independent binomial random variables.

Consider the second type of asymptotic, i.e., set the subsample sizes $n_i \to \infty$ for $i = 1, 2$. A common subsample size $n \to \infty$ is introduced to replace the individuals. Then by Classical Central Limit Theorem, the asymptotic normality of the random vector $\mathbf{Y}$ can be simply obtained as

$$\sqrt{n}(\mathbf{Y} - \mu) \xrightarrow{\mathcal{D}} N_2(0, \Sigma) \tag{2.71}$$

where $\boldsymbol{\mu} = (\pi_1, \pi_2)^T$ is the mean vector of $\mathbf{Y}$, and Σ is the asymptotic variance-covariance matrix given by

$$\Sigma = \begin{pmatrix} \pi_1(1-\pi_1) & 0 \\ 0 & \pi_2(1-\pi_2) \end{pmatrix}. \tag{2.72}$$

If the parameter vector $\boldsymbol{\beta}$ is of direct interest, therefore, it is specified by a generalized linear model with a link to a linear combination that

$$g(\boldsymbol{\mu}) = X\boldsymbol{\beta}, \tag{2.73}$$

where g satisfies the properties of a link function, X is the design matrix with specific entries. Then Delta Method may be applied appropriately to (2.71) to obtain the asymptotic normality of $\hat{\boldsymbol{\beta}}$ given by

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(0, J^{-1}(\boldsymbol{\beta})), \tag{2.74}$$

where

$$J(\boldsymbol{\beta}) = X^T W X,$$

W is the diagonal weight matrix with the elements given by

$$w_{ii} = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{[1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})]^2} \quad i = 1, 2,$$

and $\mathbf{x}_i$ is the i-th row of the design matrix.

2.5 Goodness of fit statistics

2.5.1 Deviance

One way of assessing the adequacy of a model is to compare it with a more general model, called a saturated model, with the maximum number of parameters that can be estimated. It is a generalized linear model with the same distribution and link function as the model of interest.

Let k denote the maximum number of parameters that can be estimated for the saturated model. Then k is equal to the number of potentially different linear components, which may be less than the number of observations

N. Let $\beta_{\max}$ denote the parameter vector for the saturated model and $\hat{\beta}_{\max}$ denote the maximum likelihood estimator of $\beta_{\max}$. The likelihood function for the saturated model evaluated at $\hat{\beta}_{\max}$, noted by $L(\hat{\beta}_{\max}; \mathbf{y})$, will be larger than any other likelihood function for these observations with the same assumed distribution and link function. That is because it provides the most complete description of the data. Also, denote $L(\hat{\beta}; \mathbf{y})$ the maximum value of the likelihood function for the model of interest. Therefore, the likelihood ratio

$$\lambda = \frac{L(\hat{\beta}_{\max}; \mathbf{y})}{L(\hat{\beta}; \mathbf{y})} \tag{2.75}$$

provides a way of assessing the goodness of fit for the model. In practice, the logarithm of λ, which stands for the difference between the log-likelihood functions

$$\log \lambda = l(\hat{\beta}_{\max}; \mathbf{y}) - l(\hat{\beta}; \mathbf{y})$$

is used. Large values of $\log \lambda$ suggest that the model of interest is a poor fit of the data relative to the saturated model.

In next section, the sampling distributions will be discussed. Then we may notice that $2 \log \lambda$ rather than $\log \lambda$ is the most commonly used statistic and is referred to as the deviance termed by Nelder and Wedderburn (1972). In particular, for linear logistic regression, it is given by

$$D = 2 \sum_{i=1}^{m} \left[y_i \log \left(\frac{y_i}{n_i \hat{\pi}_i} \right) + (n_i - y_i) \log \left(\frac{n_i - y_i}{n_i - n_i \hat{\pi}_i} \right) \right]. \tag{2.76}$$

2.5.2 Pearson chi-squared statistic

Instead of using maximum likelihood estimation we could estimate the parameters by minimizing the Pearson chi-squared statistic

$$X^2 = \sum \frac{(o - e)^2}{e},$$

where o represents the observed frequencies and e represents the expected frequencies. In particular, for linear logistic regression, the Pearson chi-squared statistic evaluated at the estimated expected frequencies is given by

$$X^2 = \sum_{i=1}^{m} \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}. \tag{2.77}$$

2.5.3 Equivalence

The Taylor Expansion of $s\log(s/t)$ about $s = t$ is given by

$$s\log\frac{s}{t} = (s-t) + \frac{1}{2}\frac{(s-t)^2}{t} + \cdots \quad .$$

By applying the above expansion to (2.76), we have

$$\begin{aligned} D &= 2\sum_{i=1}^{m}\Big\{(y_i - n_i\hat{\pi}_i) + \frac{1}{2}\frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i} + [(n_i - y_i) - (n_i - n_i\hat{\pi}_i)] \\ &\quad + \frac{1}{2}\frac{[(n_i - y_i) - (n_i - n_i\hat{\pi}_i)]^2}{n_i - n_i\hat{\pi}_i} + \cdots\Big\} \\ &\cong \sum_{i=1}^{m}\frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)} = X^2. \end{aligned} \tag{2.78}$$

Thus, it is that the deviance in (2.76) is asymptotically equivalent to the Pearson chi-squared statistic in (2.77).

2.6 Sampling distributions of statistics

We write the first three terms of the Taylor Expansion of the log-likelihood at $\beta = \hat{\beta}$ as

$$l(\beta) = l(\hat{\beta}) + (\beta - \hat{\beta})^T \mathrm{U}(\hat{\beta}) - \frac{1}{2}(\beta - \hat{\beta})^T \mathrm{U}'(\hat{\beta})(\beta - \hat{\beta}), \tag{2.79}$$

where $\mathrm{U}(\hat{\beta})$ is the score vector evaluated at $\beta = \hat{\beta}$ and $\mathrm{U}'(\hat{\beta})$ is the derivative of U with respect to β at $\beta = \hat{\beta}$.

Note that $\mathrm{U}(\hat{\beta}) = 0$ in (2.79) is due to the maximum likelihood estimation. If $\mathrm{U}'(\hat{\beta})$ is approximated by its expected value $\mathrm{E}(\mathrm{U}') = J$, (2.79) is therefore equal to

$$l(\beta) - l(\hat{\beta}) = -\frac{1}{2}(\beta - \hat{\beta})^T J(\hat{\beta})(\beta - \hat{\beta}), \tag{2.80}$$

where $J(\hat{\beta})$ is the information matrix evaluated at $\beta = \hat{\beta}$. Therefore, we have

$$2[l(\hat{\beta}; \mathbf{y}) - l(\beta; \mathbf{y})] = (\beta - \hat{\beta})^T J(\hat{\beta})(\beta - \hat{\beta}). \tag{2.81}$$

From the asymptotic distribution of $\hat{\beta}$, we have

$$(\beta - \hat{\beta})^T J(\hat{\beta})(\beta - \hat{\beta}) \sim \chi^2_p,$$

where χ^2_p stands for the Chi-square distribution with p degrees of freedom, and p here is also the dimensionality of β.

Thus, the sampling distribution for the deviance can be derived. Observe that

$$\begin{aligned} D &= 2[l(\hat{\beta}_{\max}; \mathbf{y}) - l(\hat{\beta}; \mathbf{y})] \\ &= 2[l(\hat{\beta}_{\max}; \mathbf{y}) - l(\beta_{\max}; \mathbf{y})] \\ &\quad -2[l(\hat{\beta}; \mathbf{y}) - l(\beta; \mathbf{y})] + 2[l(\beta_{\max}; \mathbf{y}) - l(\beta; \mathbf{y})]. \end{aligned} \tag{2.82}$$

The first term on the right-hand side of (2.82) has the χ^2_k distribution where k is the number of parameters in the saturated model. The second term has the χ^2_p distribution where p is the number of parameters in the model of interest. The third term, $v = 2[l(\beta_{\max}; \mathbf{y}) - l(\beta; \mathbf{y})]$, is a positive constant which will be near zero if the model of interest fits the data almost as well as the saturated model. Consequently, the sampling distribution of the deviance is, approximately, $\chi^2_{k-p,v}$, where v is the non-central parameter.

For logistic regression, considering the equivalence between the deviance and the Pearson chi-squared statistic, we have approximately $X^2 \sim \chi^2_{m-p}$. The choice between D and X^2 depends on the adequacy of the approximation to the χ^2_{m-p} distribution. D has a general advantage as a measure of discrepancy in that it is additive for nested sets of models if maximum likelihood estimates are used, whereas X^2 in general is not. However, there is some evidence to suggest that X^2 is often better than D because D is unduly influenced by very small frequencies (see Cressie and Read, 1989). Both of them are likely to be poor when the expected frequencies are too small.

2.7 Residual analysis

Measures of agreements between observations on a response variable and the corresponding fitted values are known as residuals. These quantities, and summary statistics derived from them, can provide much information about the adequacy of the fitted model.

For logistic regression there are two main forms of residuals corresponding to the goodness of fit measures D and X^2 respectively. Let m denote the number of observations of Y, Y_i denote the number of successes, n_i denote the number of trials in subgroups and $\hat{\pi}_i$ denote the estimated probability of being success for the ith subgroup of samples. Then the Pearson residual is defined by

$$X_i = \frac{(y_i - n_i\hat{\pi}_i)}{\sqrt{n_i\hat{\pi}_i(1-\hat{\pi}_i)}}, i = 1, \ldots, m. \tag{2.83}$$

From (2.77), $\sum_{i=1}^{m} X_i^2 = X^2$, the Pearson chi-squared goodness of fit statistic. The standardization used in the construction of the Pearson residuals does not yield residuals that have even approximate unit variance, since no allowance has been made for the inherent variation in the fitted values of the response $n_i\hat{\pi}_i$. A better procedure is to divide the raw residuals $y_i - n_i\hat{\pi}_i$ by their standard error. This standard error is quite complicated to derive, but it is found to be given by

$$\text{s.e.} = \sqrt{\hat{v}_i(1-h_i)},$$

where $\hat{v}_i = n_i\hat{\pi}_i(1-\hat{\pi}_i)$, h_i is the ith element on the diagonal of the hat matrix $H = X(X^TX)^{-1}X^T$ and the quantities h_i can be easily found through many statistical packages. So the resulting standardized residuals are

$$r_{Pi} = \frac{X_i}{\sqrt{1-h_i}}. \tag{2.84}$$

Another type of residual can be constructed from the deviance, given by

$$d_i = \text{sign}(y_i - n_i\hat{\pi}_i)\Big\{2\Big[y_i \log\Big(\frac{y_i}{n_i\hat{\pi}_i}\Big) + (n_i - y_i)\log\Big(\frac{n_i - y_i}{n_i - n_i\hat{\pi}_i}\Big)\Big]\Big\}^{\frac{1}{2}}, \tag{2.85}$$

where the term $\text{sign}(y_i - n_i\hat{\pi}_i)$ ensures that d_i has the same sign as X_i. From (2.76), $\sum_{i=1}^{m} d_i^2 = D$, the deviance. Also standardized deviance residuals are defined by

$$r_{Di} = \frac{d_i}{\sqrt{1-h_i}}. \tag{2.86}$$

These residuals can be used for checking the adequacy of a model. For instance, they should be plotted against each covariate in the model to check whether the assumption of linearity is appropriate. They should be plotted in the order of the measurements, if applicable, to check for serial correlation. Normal probability plots can also be adopted, since the standardized residuals should approximately have a standard normal distribution provided the numbers of observations for each covariate are not too small.

In the case that the data are binary or n_i is small for most covariate patterns, there are few distinct values of the residuals and, consequently, the plots may be less informative. Under this situation, the aggregated goodness of fit statistics X^2 and D may be necessary to be considered.

Sections 2.5-2.7 take Dobson (2001) for main reference.

Chapter 3

Exact simultaneous confidence bands for a simple linear regression with restricted predictor variable

In following two chapters, we consider the construction of exact hyperbolic-shape simultaneous confidence bands for a linear regression model. This chapter focuses on the construction of exact one-sided and two-sided confidence bands for a simple linear regression model with constrained predictor variable using the following three methods: the method following the idea of Bohrer (1973), the algebraical method and the tubular neighborhood method. The equivalence of the computational formulae based on these three methods is given for both one-sided and two-sided cases.

3.1 Exact one-sided confidence bands

Bohrer and Francis (1972) considered an one-sided confidence bound of hyperbolic shape for a simple linear regression model

$$y_i = f_i(x;\beta) + \varepsilon_i \tag{3.1}$$

with

$$f_i(x;\boldsymbol{\beta}) = \beta_1 + \beta_2(x_i - \bar{x}), i = 1, \ldots, n,$$

where y_i's are the observations of the response, the differences between the observations of the only predictor variable and their mean value $(x_i - \bar{x})$'s are restricted in a given interval $[a, b]$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ is the vector of unknown regression coefficients, ε_i's are independent and identically distributed normal random errors with mean 0 and unknown variance σ^2. If we define $S_x = \sum(x_i - \bar{x})^2$ and $S_{xy} = \sum(x_i - \bar{x})y_i$, then the least squares estimator of $\boldsymbol{\beta}$ and the usual unbiased estimator of σ^2 are given by $\hat{\boldsymbol{\beta}} = (\sum y_i/n, S_{xy}/S_x)^T$ and $\hat{\sigma}^2 = \sum(y_i - \hat{\beta}_1 - \hat{\beta}_2(x_i - \bar{x}))/(n-2)$ respectively, which are independent by studying least squares theory and have the following distributions

$$\hat{\boldsymbol{\beta}} \sim N_2\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \begin{pmatrix} n^{-1}\sigma^2 & 0 \\ 0 & S_x^{-1}\sigma^2 \end{pmatrix}\right),$$

$$(n-2)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-2}.$$

An one-sided hyperbolic-shape simultaneous confidence band for the mean responses

$$f(x;\boldsymbol{\beta}) = \beta_1 + \beta_2(x - \bar{x})$$

is centered by $f(x;\hat{\boldsymbol{\beta}})$ and with band width proportional to the standard deviation of $f(x;\hat{\boldsymbol{\beta}})$. Specifically, the band, e.g., with upper bound, is given by

$$f(x;\boldsymbol{\beta}) \leq f(x;\hat{\boldsymbol{\beta}}) + c\hat{\sigma}H(x;\hat{\boldsymbol{\beta}}), \text{ for all } x - \bar{x} \in [a, b],$$

where c is a critical value and

$$H(x;\hat{\boldsymbol{\beta}}) = [\mathrm{Var} f(x;\hat{\boldsymbol{\beta}})]^{\frac{1}{2}}/\sigma = n^{-1} + S_x^{-1}(x - \bar{x})^2.$$

The key of constructing a simultaneous confidence band is to find an appropriate critical value c such that the band has the coverage probability defined by

$$\mathrm{P}(c) = \mathrm{P}\{f(x;\boldsymbol{\beta}) \leq f(x;\hat{\boldsymbol{\beta}}) + c\hat{\sigma}H(x;\hat{\boldsymbol{\beta}}), x - \bar{x} \in [a, b]\} \tag{3.2}$$

equal to a preassigned confidence level $1 - \alpha$.

Let $\mathbf{z} = (n^{-1/2}, (x-\bar{x})S_x^{-1/2})^T$ and $\mathbf{N} = ((\beta_1 - \hat{\beta}_1)n^{1/2}, (\beta_2 - \hat{\beta}_2)S_x^{1/2})^T$. Note that since $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)^T$ is independent of $\hat{\sigma}$, so is $\mathbf{N}$. Let $\mathbf{t} = \mathbf{N}/\hat{\sigma}$, then we have

$$\begin{aligned}
& \beta_1 + \beta_2(x-\bar{x}) \le \hat{\beta}_1 + \hat{\beta}_2(x-\bar{x}) + c\hat{\sigma}[n^{-1} + S_x^{-1}(x-\bar{x})^2] \\
\Leftrightarrow\ & (\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)(x-\bar{x}) \le c\hat{\sigma}[n^{-1} + S_x^{-1}(x-\bar{x})^2] \\
\Leftrightarrow\ & (n^{-\frac{1}{2}}, (x-\bar{x})S_x^{-\frac{1}{2}})^T \left[((\beta_1 - \hat{\beta}_1)n^{\frac{1}{2}}, (\beta_2 - \hat{\beta}_2)S_x^{\frac{1}{2}})/\hat{\sigma} \right] \\
& \le c \| (n^{-\frac{1}{2}}, (x-\bar{x})S_x^{-\frac{1}{2}}) \|.
\end{aligned}$$

Consequently, the confidence level of the band in (3.2) is equal to

$$\mathrm{P}(c) = \mathrm{P}\{\mathbf{z}^T\mathbf{t} \le c\|\mathbf{z}\|, \text{ for } x - \bar{x} \in [a,b]\}, \tag{3.3}$$

where $x - \bar{x} \in [a,b]$ determines a restricted region for $\mathbf{z}$ in terms of a and b. From the definition of $\mathbf{z}$, it is clear that $\mathbf{z}$ has the fixed first coordinate and the second coordinate bounded by an interval as $S_x^{-1/2}$ is known. This implies that $\mathbf{z}$ varies within a circular cone $\mathcal{Z} = \{\mathbf{z} : z_1 \ge q\|\mathbf{z}\|\}$, where z_1 is the first coordinate of $\mathbf{z}$ and q is a constant which will be explicitly given in the following text when needed. Therefore, $\mathrm{P}(c)$ is equal to $\mathrm{P}\{\mathbf{t} \in R\}$, where

$$R = \{\mathbf{t} : \mathbf{z}^T\mathbf{t} \le c\|\mathbf{z}\|, \text{ all } \mathbf{z} \in \mathcal{Z}\}. \tag{3.4}$$

This is the starting point of the following three methods we are going to discuss.

3.1.1 Method following the idea of Bohrer (1973)

Let $\mathbf{a} = (n^{-1/2}, aS_x^{-1/2})^T$ and $\mathbf{b} = (n^{-1/2}, bS_x^{-1/2})^T$ be the boundaries of $\mathcal{Z}$, and $\phi^* \in [0, \pi]$ be the angle between $\mathbf{a}$ and $\mathbf{b}$. Set up a coordinates system such that the horizontal axis has the same direction as $\mathbf{a}$. Let $\phi_{\mathbf{t}}$ be the angle of $\mathbf{t}$ turned moving anti-clockwise from $\mathbf{a}$ to $\mathbf{t}$.

Lemma 3.1.1 *Under the notations of* $\mathbf{t}$*,* ϕ^* *and* $\phi_{\mathbf{t}}$*,* R *in (3.4) can be par-*

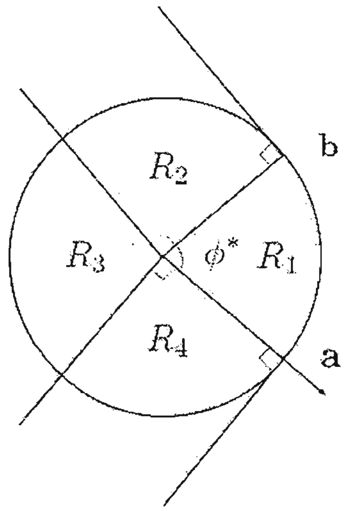

Figure 3.1: For the method following Bohrer (1973) in one-sided case

titioned into four disjoint parts according to the location of t*:*

$$R_1 = \{\mathbf{t} : \|\mathbf{t}\| \leq c, 0 < \phi_{\mathbf{t}} < \phi^*\}, \tag{3.5}$$

$$R_2 = \{\mathbf{t} : 0 \leq \mathbf{b}^T\mathbf{t} \leq c\|\mathbf{b}\|, \phi^* \leq \phi_{\mathbf{t}} < \phi^* + \frac{1}{2}\pi\}, \tag{3.6}$$

$$R_3 = \{\mathbf{t} : \phi^* + \frac{1}{2}\pi \leq \phi_{\mathbf{t}} < \frac{3}{2}\pi\}, \tag{3.7}$$

$$R_4 = \{\mathbf{t} : 0 \leq \mathbf{a}^T\mathbf{t} \leq c\|\mathbf{a}\|, \frac{3}{2}\pi \leq \phi_{\mathbf{t}} \leq 2\pi\}. \tag{3.8}$$

Proof. When $\mathbf{t} \in R_1$, then $\mathbf{t} \in \mathcal{Z}$. We have $\|\mathbf{t}\|^2 \leq c\|\mathbf{t}\|$ from (3.5), and further $\mathbf{t}^T\mathbf{t} \leq c\|\mathbf{t}\|$ which implies $\mathbf{t} \in R$ by studying (3.4). When $\mathbf{t} \in R_2$, since $\mathbf{b} \in \mathcal{Z}$, obviously $\mathbf{t} \in R$. Similarly, when $\mathbf{t} \in R_4$, $\mathbf{t} \in R$. Finally when $\mathbf{t} \in R_3$, since $\mathbf{z} \in \mathcal{Z}$, we have $\pi/2 \leq \phi_{\mathbf{t}} - \phi_{\mathbf{z}} \leq 3\pi/2$. Hence $\mathbf{z}^T\mathbf{t} \leq 0$ which implies $\mathbf{t} \in R$. Therefore, $\cup_{i=1}^4 R_i \subset R$.

Conversely, when $\mathbf{t} \in R$ and $0 < \phi_{\mathbf{t}} < \phi^*$, we have $\mathbf{t} \in \mathcal{Z}$ from the definition of ϕ^*. Then $\mathbf{t}^T\mathbf{t} \leq c\|\mathbf{t}\|$ implies $\|\mathbf{t}\| \leq c$ which is equivalent to $\mathbf{t} \in R_1$. Consider $\mathbf{t} \in R$ and $\phi^* \leq \phi_{\mathbf{t}} < \phi^* + \pi/2$, only $\mathbf{b} \in \mathcal{Z}$ in this case, so we have $\mathbf{b}^T\mathbf{t} \leq c\|\mathbf{b}\|$ which implies $\mathbf{t} \in R_2$. Similarly, when $\mathbf{t} \in R$ and $3\pi/2 \leq \phi_{\mathbf{t}} \leq 2\pi$, we have $\mathbf{a}^T\mathbf{t} \leq c\|\mathbf{a}\|$ which implies $\mathbf{t} \in R_4$. As for the case when $\mathbf{t} \in R$ and $\phi^* + \pi/2 \leq \phi_{\mathbf{t}} < 3\pi/2$, $\mathbf{z}^T\mathbf{t} \leq 0$ for all $\mathbf{z} \in \mathcal{Z}$, we therefore have $\mathbf{t} \in R_3$. Consequently, $R \subset \cup_{i=1}^4 R_i$. Overall, $\cup_{i=1}^4 R_i = R$. #

By applying Lemma 3.1.1 to $\mathrm{P}\{\mathbf{t} \in R\}$ with R defined in (3.4), we have the confidence level of the band based on this method equal to

$$\mathrm{P}_B(c) = \sum_{i=1}^{4} \mathrm{P}\{\mathbf{t} \in R_i\}, \tag{3.9}$$

where the four individual probabilities on the right-hand side of (3.9) can be evaluated separately. Define the polar coordinates of $\mathbf{t}$ in terms of $(R_{\mathbf{t}}, \phi_{\mathbf{t}})$ that $\mathbf{t} = (R_{\mathbf{t}} \cos\phi_{\mathbf{t}}, R_{\mathbf{t}} \sin\phi_{\mathbf{t}})$. Note that $\mathbf{t}$ can be written in terms of the polar coordinates of $\mathbf{N}$ as $((R_{\mathbf{N}}/\hat{\sigma})\cos\phi_{\mathbf{N}}, (R_{\mathbf{N}}/\hat{\sigma})\sin\phi_{\mathbf{N}})$. Note that $\phi_{\mathbf{N}}$ and $\phi_{\mathbf{t}}$ denote the same angle because $\mathbf{N}/\hat{\sigma}$ does not change the location of $\mathbf{N}$. As we know that $\mathbf{N}$ has a bivariate standard normal distribution, one may find the joint density function of $R_{\mathbf{N}}$ and $\phi_{\mathbf{N}}$ via the transformation of random variables. By finding the individual marginal density functions of $R_{\mathbf{N}}$ and $\phi_{\mathbf{N}}$, we have that the joint density is equal to the product of the individual marginal densities. And this implies that $R_{\mathbf{N}}$ is independent of $\phi_{\mathbf{N}}$. Accordingly, $R_{\mathbf{t}}$ is independent of $\phi_{\mathbf{t}}$. In addition, $\|\mathbf{t}\|^2/2 = (\|\mathbf{N}\|^2/2)/\hat{\sigma}^2 = (\|\mathbf{N}/\sigma\|^2/2)/(\hat{\sigma}^2/\sigma^2)$ has the $F_{2,\nu}$ distribution, and $\phi_{\mathbf{t}}$ has the uniform marginal distribution.

Now, we turn to evaluate the probabilities on the right-hand side of (3.9) individually. Specifically, we have

$$\begin{aligned}
\mathrm{P}\{\mathbf{t} \in R_1\} &= \mathrm{P}\{\|\mathbf{t}\| \leq c, 0 < \phi_{\mathbf{t}} < \phi^*\} \\
&= \mathrm{P}\{\|\mathbf{t}\| \leq c\} \cdot \mathrm{P}\{0 < \phi_{\mathbf{t}} < \phi^*\} \\
&= \mathrm{P}\{\|\mathbf{t}\|^2/2 \leq c^2/2\} \cdot \frac{\phi^*}{2\pi} \\
&= \frac{\phi^*}{2\pi} F_{2,\nu}(\frac{c^2}{2}), && (3.10) \\
\mathrm{P}\{\mathbf{t} \in R_3\} &= \mathrm{P}\{\phi^* + \frac{1}{2}\pi \leq \phi_{\mathbf{t}} < \frac{3}{2}\pi\} \\
&= [\frac{3}{2}\pi - (\phi^* + \frac{1}{2}\pi)]/2\pi \\
&= \frac{1}{2} - \frac{\phi^*}{2\pi}, && (3.11)
\end{aligned}$$

where $F_{2,\nu}$ stands for the F cumulative distribution function with 2 and ν degrees of freedom. As for the other two probabilities, take the case of $\mathbf{t} \in R_2$ for example. From (3.6), $\mathbf{b}^T\mathbf{t} \le c\|\mathbf{b}\|$ implies $\|\mathbf{t}\|\cos[\phi_{\mathbf{t}} - \phi(\mathbf{b})] \le c$. If we rotate the coordinates system such that the horizontal axis has the same direction as $\mathbf{b}$, then $\|\mathbf{t}\|\cos[\phi_{\mathbf{t}} - \phi(\mathbf{b})]$ can be thought as the projection of $\mathbf{t}$ on the horizontal axis, namely, the first coordinate of $\mathbf{t}$. Doing this does not change the probability. Thus we have

$$\begin{aligned} \mathrm{P}\{\mathbf{t} \in R_2\} &= \mathrm{P}\{0 < t_1 \le c\} \cdot \mathrm{P}\{\phi^* \le \phi_{\mathbf{t}} < \phi^* + \frac{\pi}{2}\} \\ &= \mathrm{P}\{\|t_1\|^2 \le c^2\} \cdot (\phi^* + \frac{\pi}{2} - \phi^*)/2\pi \\ &= \frac{1}{4} F_{1,\nu}(c^2). \end{aligned} \tag{3.12}$$

Similarly,

$$\mathrm{P}\{\mathbf{t} \in R_4\} = \frac{1}{4} F_{1,\nu}(c^2).$$

Overall,

$$\mathrm{P}_B(c) = \frac{\phi^*}{2\pi} F_{2,\nu}(\frac{c^2}{2}) + \frac{1}{2} F_{1,\nu}(c^2) + (\frac{1}{2} - \frac{\phi^*}{2\pi}). \tag{3.13}$$

3.1.2 Algebraical method

From (3.3) and (3.4), the confidence level can be alternatively written as

$$\mathrm{P}(c) = \mathrm{P}\Big\{ \sup_{\mathbf{z} \in \mathcal{Z}} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|} \le c \Big\}. \tag{3.14}$$

The key idea of the algebraical method is to find the explicit form of the supreme in (3.14).

Lemma 3.1.2 *Rotating the coordinates such that the horizontal axis reaches the central direction of $\mathcal{Z}$. Let $\theta^* = (1/2)\phi^*$ be the angle between the horizontal axis and one of the boundaries of $\mathcal{Z}$, say, $\mathbf{b}$. Also, let $\theta_{\mathbf{t}}$ denote the angle between the horizontal axis and $\mathbf{t}$. Then we have*

$$\sup_{\mathbf{z} \in \mathcal{Z}} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|} = \begin{cases} \|\mathbf{t}\| & \text{if } \quad -\theta^* \le \theta_{\mathbf{t}} \le \theta^*, \\ \|\mathbf{t}\|\cos(\theta_{\mathbf{t}} - \theta^*) & \text{if } \quad \theta^* < \theta_{\mathbf{t}} < \pi, \\ \|\mathbf{t}\|\cos(2\pi - \theta_{\mathbf{t}} - \theta^*) & \text{if } \quad \pi < \theta_{\mathbf{t}} < 2\pi - \theta^*. \end{cases}$$

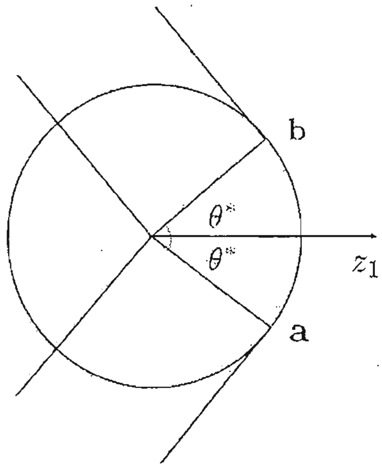

Figure 3.2: For the algebraical method in one-sided case

Proof. When $-\theta^* \le \theta_{\mathrm{t}} \le \theta^*$, $\mathrm{t} \in \mathcal{Z}$ which implies the supreme is equal to $\mathbf{t}^T\mathbf{t}/\|\mathbf{t}\| = \|\mathbf{t}\|$. When $\theta^* < \theta_{\mathrm{t}} < \pi$, the supreme is equal to $\|\mathbf{t}\|\cos(\theta_{\mathrm{t}} - \theta^*)$ since $\cos\theta$ decreases with θ. Similarly, the supreme is equal to $\|\mathbf{t}\|\cos(2\pi - \theta_{\mathrm{t}} - \theta^*)$ when $\pi < \theta_{\mathrm{t}} < 2\pi - \theta^*$. #

Applying Lemma 3.1.2, the confidence level (3.14) based on the algebraical method becomes

$$\begin{aligned}
\mathrm{P}_A(c) &= \mathrm{P}\Big\{\sup_{\mathbf{z}\in\mathcal{Z}} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|} \le c\Big\} \\
&= \mathrm{P}\{\|\mathbf{t}\| \le c, -\theta^* \le \theta_{\mathrm{t}} \le \theta^*\} \\
&\quad + \mathrm{P}\{\|\mathbf{t}\|\cos(\theta_{\mathrm{t}} - \theta^*) \le c, \theta^* < \theta_{\mathrm{t}} < \pi\} \\
&\quad + \mathrm{P}\{\|\mathbf{t}\|\cos(2\pi - \theta_{\mathrm{t}} - \theta^*) \le c, \pi < \theta_{\mathrm{t}} < 2\pi - \theta^*\}. \quad (3.15)
\end{aligned}$$

Note that the second and the third terms on the right-hand side of (3.15) are in fact the same because the regions of $\mathbf{t}$ corresponding to these two cases are graphically symmetric. Therefore, (3.15) is further equal to

$$
\begin{aligned}
\mathrm{P}_A(c) &= \mathrm{P}\{\|\mathrm{t}\| \le c, -\theta^* \le \theta_{\mathrm{t}} \le \theta^*\} \\
&\quad + 2\mathrm{P}\{\|\mathrm{t}\|\cos(\theta_{\mathrm{t}} - \theta^*) \le c, \theta^* < \theta_{\mathrm{t}} < \pi\} \\
&= \mathrm{P}\{\|\mathrm{t}\| \le c, -\theta^* \le \theta_{\mathrm{t}} \le \theta^*\} \\
&\quad + 2\Big(\mathrm{P}\{0 \le \|\mathrm{t}\|\cos(\theta_{\mathrm{t}} - \theta^*) \le c, \theta^* < \theta_{\mathrm{t}} < \theta^* + \frac{\pi}{2}\} \\
&\quad + \mathrm{P}\{\|\mathrm{t}\|\cos(\theta_{\mathrm{t}} - \theta^*) < 0, \theta^* + \frac{\pi}{2} < \theta_{\mathrm{t}} < \pi\}\Big) \\
&= \mathrm{P}\{\frac{\|\mathrm{t}\|^2}{2} \le \frac{c^2}{2}\} \cdot \mathrm{P}\{-\theta^* \le \theta_{\mathrm{t}} \le \theta^*\} \\
&\quad + 2\Big(\mathrm{P}\{\frac{\|\mathrm{t}\|^2}{2} \le \frac{c^2}{2\cos^2(\theta_{\mathrm{t}} - \theta^*)}\} \cdot \mathrm{P}\{\theta^* < \theta_{\mathrm{t}} < \theta^* + \frac{\pi}{2}\} \\
&\quad + \mathrm{P}\{\theta^* + \frac{\pi}{2} < \theta_{\mathrm{t}} < \pi\}\Big) \\
&= \frac{2\theta^*}{2\pi} F_{2,\nu}(\frac{c^2}{2}) + 2\Big(\int_{\theta^*}^{\theta^* + \frac{\pi}{2}} \frac{1}{2\pi} F_{2,\nu}\Big(\frac{c^2}{2\cos^2(\theta - \theta^*)}\Big) d\theta + \frac{\pi/2 - \theta^*}{2\pi}\Big) \\
&= \frac{\theta^*}{\pi} F_{2,\nu}(\frac{c^2}{2}) + \frac{1}{\pi}\int_0^{\frac{\pi}{2}} F_{2,\nu}(\frac{c^2}{2\cos^2\theta}) d\theta + (\frac{1}{2} - \frac{\theta^*}{\pi}), \qquad (3.16)
\end{aligned}
$$

where $F_{2,\nu}$ denotes the cumulative distribution function of $F_{2,\nu}$ distribution.

3.1.3 Tubular neighborhood method

The idea of this method seems similar to the thoughts in Naiman (1986, 1990), Sun and Loader (1994). Here, the exact volume of the tubular neighborhood of a circular cone is calculated to evaluate the coverage probability of the one-sided confidence band.

From (3.14), the confidence level is given by the alternative form

$$
\mathrm{P}(c) = \mathrm{P}\Big\{\sup_{\mathrm{z} \in \mathcal{Z}} \frac{\mathrm{z}^T\mathrm{t}}{\|\mathrm{z}\|\|\mathrm{t}\|} \le \frac{c}{\|\mathrm{t}\|}\Big\}. \qquad (3.17)
$$

Note that $\mathrm{t}/\|\mathrm{t}\|$ is independent of $\|\mathrm{t}\|$ and so is $c/\|\mathrm{t}\|$. And the supreme in

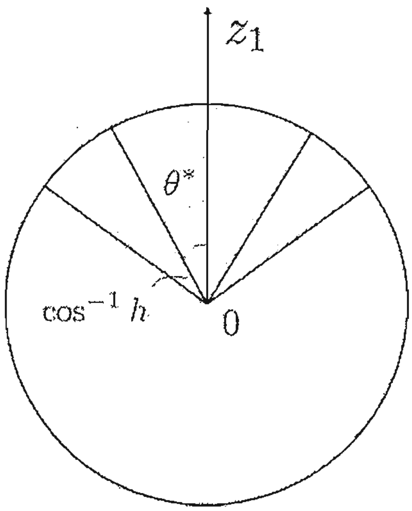

Figure 3.3: For the tubular neighborhood method in one-sided case

(3.17) is no larger than one. Then (3.17) is further equal to

$$\begin{aligned}
& 1 - \mathrm{P}\Big\{\sup_{\mathbf{z}\in\mathcal{Z}} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|\|\mathbf{t}\|} > \frac{c}{\|\mathbf{t}\|}\Big\} \\
= \; & 1 - \int_0^{\infty} \mathrm{P}\Big\{\sup_{\mathbf{z}\in\mathcal{Z}} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|\|\mathbf{t}\|} > \frac{c}{\sqrt{2w}}\Big\} \cdot dF_{2,\nu}(w) \\
= \; & 1 - \int_{\frac{c^2}{2}}^{\infty} \mathrm{P}\Big\{\sup_{\mathbf{z}\in\mathcal{Z}} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|\|\mathbf{t}\|} > \frac{c}{\sqrt{2w}}\Big\} \cdot dF_{2,\nu}(w). \qquad (3.18)
\end{aligned}$$

Let $0 < h = c/\sqrt{2w} < 1$. The set

$$\mathrm{E}(h) = \Big\{\mathbf{t} : \sup_{\mathbf{z}\in\mathcal{Z}} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|\|\mathbf{t}\|} > h\Big\}$$

contains all possible $\mathbf{t}$'s with the angle between $\mathbf{t}$ and $\mathbf{z}$ being at largest $\cos^{-1} h$. So $\mathrm{E}(h)$ is in fact a fan with vertex at the origin, symmetrically containing the region of $\mathcal{Z}$, and has the angle $\cos^{-1} h$ between one of its bound and the nearest bound of $\mathcal{Z}$. Therefore, $\mathrm{P}\{\mathbf{t} \in \mathrm{E}(h)\}$ is equal to

$2\mathrm{P}\{0 \le \theta_{\mathbf{t}} \le \theta^* + \cos^{-1} h\}$. Consequently, (3.18) becomes

$$
\begin{aligned}
& 1 - \int_{\frac{c^2}{2}}^{\infty} 2 \cdot \frac{\theta^* + \cos^{-1}(c/\sqrt{2w})}{2\pi} \cdot dF_{2,\nu}(w) \\
= \; & 1 - \int_{\frac{c^2}{2}}^{\infty} \frac{\theta^* + \cos^{-1}(c/\sqrt{2w})}{\pi} \cdot dF_{2,\nu}(w). \qquad (3.19)
\end{aligned}
$$

3.1.4 Equivalence of the formulae

It is of natural interest to compare the three computational formulae (3.13), (3.16) and (3.19) corresponding to the three methods respectively. Formula (3.13) comes from the original paper of Bohrer and Francis (1972). We first derive the equivalent formula of (3.13) by rotating the coordinates system.

Rotating the coordinates system such that the central direction of $\mathcal{Z}$ is given by z_1 axis. In this case, by defining $\theta^* = \phi^*/2$ and $\theta_{\mathbf{t}}$ as the angle between z_1 axis and the vector $\mathbf{t}$, the confidence level based on the first method is then $\mathrm{P}_B(c) = \sum_{i=1}^{4} \mathrm{P}\{\mathbf{t} \in R_i'\}$, where

$$
\begin{aligned}
R_1' &= \{\mathbf{t} : \|\mathbf{t}\| \le c, -\theta^* < \theta_{\mathbf{t}} < \theta^*\}, \\
R_2' &= \{\mathbf{t} : \|\mathbf{t}\| \cos(\theta_{\mathbf{t}} - \theta^*) \le c, \theta^* < \theta_{\mathbf{t}} < \theta^* + \frac{\pi}{2}\}, \\
R_3' &= \{\mathbf{t} : \theta^* + \frac{\pi}{2} < \theta_{\mathbf{t}} < \frac{3\pi}{2} - \theta^*\}, \\
R_4' &= \{\mathbf{t} : \|\mathbf{t}\| \cos(\theta_{\mathbf{t}} - \theta^*) \le c, \frac{3\pi}{2} - \theta^* < \theta_{\mathbf{t}} < 2\pi\}. \qquad (3.20)
\end{aligned}
$$

So we have

$$
\begin{aligned}
\mathrm{P}_B(c) &= \int_{-\theta^*}^{\theta^*} \frac{1}{2\pi} \mathrm{P}\{\|\mathbf{t}\| \le c\} d\theta + 2 \int_{\theta^*}^{\theta^* + \frac{\pi}{2}} \frac{1}{2\pi} \mathrm{P}\{\|\mathbf{t}\| \le \frac{c}{\cos(\theta - \theta^*)}\} d\theta \\
&\quad + \int_{\theta^* + \frac{\pi}{2}}^{\frac{3\pi}{2} - \theta^*} \frac{1}{2\pi} d\theta \\
&= \frac{\theta^*}{\pi} F_{2,\nu}(\frac{c^2}{2}) + \frac{1}{\pi} \int_0^{\frac{\pi}{2}} F_{2,\nu}(\frac{c^2}{2\cos^2\theta}) d\theta + \frac{\pi - 2\theta^*}{2\pi} \qquad (3.21) \\
&= \frac{\theta^*}{\pi} F_{2,\nu}(\frac{c^2}{2}) + \frac{1}{\pi} \int_0^{\frac{\pi}{2}} d\theta \int_0^{\frac{c^2}{2\cos^2\theta}} dF_{2,\nu}(w) + \frac{\pi - 2\theta^*}{2\pi}. \qquad (3.22)
\end{aligned}
$$

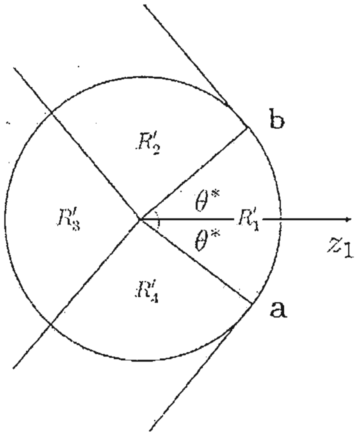

Figure 3.4: Picture obtained by rotating the coordinates system

Note that (3.21) is the same as (3.16) which implies that the method following the idea of Bohrer (1973) can have the same formula as the algebraical method. The second term of (3.22) is further equal to

$$\begin{aligned} & \frac{1}{\pi}\Big(\int_0^{\frac{\pi}{2}} d\theta \int_0^{\frac{c^2}{2}} dF_{2,\nu}(w) + \int_0^{\frac{\pi}{2}} d\theta \int_{\frac{c^2}{2}}^{\frac{c^2}{2\cos^2\theta}} dF_{2,\nu}(w)\Big) \\ = \; & \frac{1}{2}F_{2,\nu}(\frac{c^2}{2}) + \frac{1}{\pi}\int_{\frac{c^2}{2}}^{\infty}\int_{\cos^{-1}(\frac{c}{\sqrt{2w}})}^{\frac{\pi}{2}} d\theta \cdot dF_{2,\nu}(w). \end{aligned} \tag{3.23}$$

By substituting (3.23) into (3.22), we have

$$\begin{aligned} \mathrm{P}_B(c) = \; & (\frac{\theta^*}{\pi} + \frac{1}{2})F_{2,\nu}(\frac{c^2}{2}) + \frac{\pi - 2\theta^*}{2\pi} + \\ & \frac{1}{\pi}\int_{\frac{c^2}{2}}^{\infty}\int_{\cos^{-1}(\frac{c}{\sqrt{2w}})}^{\frac{\pi}{2}} d\theta \cdot dF_{2,\nu}(w). \end{aligned} \tag{3.24}$$

Now consider formula (3.19). It can be further written as

$$
\begin{aligned}
& 1-\int_{\frac{c^2}{2}}^{\infty}\frac{\theta^*}{\pi}dF_{2,\nu}(w)-\int_{\frac{c^2}{2}}^{\infty}\frac{\cos^{-1}(c/\sqrt{2w})}{\pi}dF_{2,\nu}(w) \\
= & 1-\frac{\theta^*}{\pi}F_{2,\nu}(w>\frac{c^2}{2})-\frac{1}{\pi}\int_{\frac{c^2}{2}}^{\infty}\int_{0}^{\cos^{-1}(\frac{c}{\sqrt{2w}})}d\theta\cdot dF_{2,\nu}(w) \\
= & \frac{\theta^*}{\pi}F_{2,\nu}(\frac{c^2}{2})+\frac{\pi-\theta^*}{\pi}-\frac{1}{\pi}\int_{\frac{c^2}{2}}^{\infty}\Big(\int_{0}^{\frac{\pi}{2}}-\int_{\cos^{-1}(\frac{c}{\sqrt{2w}})}^{\frac{\pi}{2}}\Big)d\theta\cdot dF_{2,\nu}(w) \\
= & \frac{\theta^*}{\pi}F_{2,\nu}(\frac{c^2}{2})+\frac{\pi-\theta^*}{\pi}-\frac{1}{2}F_{2,\nu}(w>\frac{c^2}{2})+\frac{1}{\pi}\int_{\frac{c^2}{2}}^{\infty}\int_{\cos^{-1}(\frac{c}{\sqrt{2w}})}^{\frac{\pi}{2}}d\theta\cdot dF_{2,\nu}(w) \\
= & (\frac{\theta^*}{\pi}+\frac{1}{2})F_{2,\nu}(\frac{c^2}{2})+\frac{\pi-2\theta^*}{2\pi}+\frac{1}{\pi}\int_{\frac{c^2}{2}}^{\infty}\int_{\cos^{-1}(\frac{c}{\sqrt{2w}})}^{\frac{\pi}{2}}d\theta\cdot dF_{2,\nu}(w), \qquad (3.25)
\end{aligned}
$$

which is equivalent to (3.24).

Hence, it can be concluded that the three methods of constructing one-sided confidence bands for a simple linear regression model give the same result mathematically. Clearly, (3.13) is relatively simple compared with (3.16) and (3.19) since both (3.16) and (3.19) involve an integration.

3.2 Exact two-sided confidence bands

It is also of interest to think about constructing exact two-sided simultaneous confidence bands for a simple linear regression. Recall (3.4), we have the following setting corresponding to the two-sided case that

$$R=\{\mathbf{t}:|\mathbf{z}^T\mathbf{t}|\le c\|\mathbf{z}\|,\ \text{all } \mathbf{z}\in\mathcal{Z}\}, \qquad (3.26)$$

where $\mathbf{t}=\mathbf{N}/\hat{\sigma}$, $\mathbf{N}$ has the $N_2(0,\sigma^2 I)$ distribution, $\hat{\sigma}$ is the usual unbiased estimator of unknown σ and has the $\sigma\sqrt{\chi^2_\nu/\nu}$ distribution with ν degrees of freedom, c is a critical value, and

$$\mathcal{Z}=\{\mathbf{z}:|z_1|\ge q\|\mathbf{z}\|\}, \qquad (3.27)$$

where q is a non-negative constant.

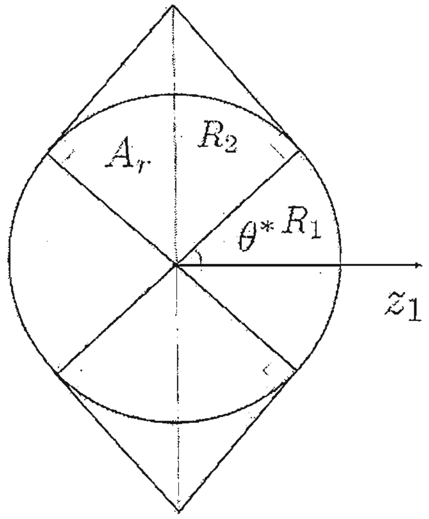

Figure 3.5: For the method following Bohrer (1973) in two-sided case

3.2.1 Method following the idea of Bohrer (1973)

R is shown in Figure 3.5. Since R has a symmetric structure, the probability $\mathrm{P}\{\mathbf{t} \in R\}$ equals four times the summation of the probabilities $\mathrm{P}\{\mathbf{t} \in R_1\}$ and $\mathrm{P}\{\mathbf{t} \in R_2\}$, where

$$\begin{aligned} \theta^* &= \cos^{-1} q, \\ R_1 &= \{\mathbf{t} : \|\mathbf{t}\| \le c, 0 < \theta_{\mathbf{t}} < \theta^*\}, \\ R_2 &= \{\mathbf{t} : \|\mathbf{t}\| \cos(\theta_{\mathbf{t}} - \theta^*) \le c, \theta^* < \theta_{\mathbf{t}} < \frac{\pi}{2}\}, \end{aligned} \tag{3.28}$$

$\theta_{\mathbf{t}}$ is the angle between $\mathbf{t}$ and the z_1 axis.

Note that $\|\mathbf{N}\|$ is independent of $\hat{\sigma}$ by studying the least squares theory and $\|\mathbf{t}\|^2/2$ has the F distribution with 2 and ν degrees of freedom. Also, note that $\|\mathbf{t}\|$ is independent of $\theta_{\mathbf{t}}$, which has been shown previously. Therefore, we have

$$\begin{aligned}
& \mathrm{P}_B\{\mathbf{t} \in R\} \\
= \;& 4\Big(\mathrm{P}\{\mathbf{t} \in R_1\} + \mathrm{P}\{\mathbf{t} \in R_2\}\Big) \\
= \;& 4\Big(\mathrm{P}\{\|\mathbf{t}\| \le c, 0 < \theta_{\mathbf{t}} < \theta^*\} \\
& +\mathrm{P}\{\|\mathbf{t}\| \cos(\theta_{\mathbf{t}} - \theta^*) \le c, \theta^* < \theta_{\mathbf{t}} < \frac{\pi}{2}\}\Big) \\
= \;& 4\Big(\frac{\theta^*}{2\pi} F_{2,\nu}(\frac{c^2}{2}) + \int_{\theta^*}^{\frac{\pi}{2}} \frac{1}{2\pi} F_{2,\nu}(\frac{c^2}{2\cos^2(\theta - \theta^*)}) d\theta\Big) \\
= \;& \frac{2\theta^*}{\pi} F_{2,\nu}(\frac{c^2}{2}) + \frac{2}{\pi} \int_0^{\frac{\pi}{2}-\theta^*} F_{2,\nu}(\frac{c^2}{2\cos^2\theta}) d\theta. \qquad (3.29)
\end{aligned}$$

Note that this method agrees with that of Wynn and Bloomfield (1971).

3.2.2 Algebraical method

According to (3.26) and (3.27), R has the alternative form

$$R = \Big\{\mathbf{t} : \sup_{\mathbf{z} \in \mathcal{Z}} \frac{|\mathbf{z}^T \mathbf{t}|}{\|\mathbf{z}\|} \le c\Big\}, \qquad (3.30)$$

where the supreme can be found directly and explicitly.

Lemma 3.2.1 *Under the notations of $\theta_{\mathbf{t}}$ and θ^*, we have*

$$\sup_{\mathbf{z} \in \mathcal{Z}} \frac{|\mathbf{z}^T \mathbf{t}|}{\|\mathbf{z}\|} = \begin{cases} \|\mathbf{t}\| & \text{if } \theta_{\mathbf{t}} \in [0, \theta^*] \cup [\pi - \theta^*, \pi + \theta^*] \\ & \quad \cup [2\pi - \theta^*, 2\pi], \\ \|\mathbf{t}\| |\cos(\theta_{\mathbf{t}} - \theta^*)| & \text{if } \theta_{\mathbf{t}} \in [\theta^*, \frac{\pi}{2}] \cup [\pi + \theta^*, \frac{3\pi}{2}], \\ \|\mathbf{t}\| |\cos(\pi - \theta_{\mathbf{t}} - \theta^*)| & \text{if } \theta_{\mathbf{t}} \in [\frac{\pi}{2}, \pi - \theta^*] \cup [\frac{3\pi}{2}, 2\pi - \theta^*]. \end{cases}$$

The proof of Lemma 3.2.1 is very similar with that of Lemma 3.1.2. #

By applying Lemma 3.2.1, we have

$$
\begin{aligned}
& \mathrm{P}_A\{\mathbf{t} \in R\} \\
= \;& 2 \cdot \frac{2\theta^*}{2\pi} \mathrm{P}\{\|\mathbf{t}\| \le c\} \\
& + \int_{\theta^*}^{\frac{\pi}{2}} \frac{1}{2\pi} \mathrm{P}\{\|\mathbf{t}\| |\cos(\theta - \theta^*)| \le c\} d\theta \\
& + \int_{\pi+\theta^*}^{\frac{3\pi}{2}} \frac{1}{2\pi} \mathrm{P}\{\|\mathbf{t}\| |\cos(\theta - \pi - \theta^*)| \le c\} d\theta \\
& + \int_{\frac{\pi}{2}}^{\pi-\theta^*} \frac{1}{2\pi} \mathrm{P}\{\|\mathbf{t}\| |\cos(\pi - \theta - \theta^*)| \le c\} d\theta \\
& + \int_{\frac{3\pi}{2}}^{2\pi-\theta^*} \frac{1}{2\pi} \mathrm{P}\{\|\mathbf{t}\| |\cos(2\pi - \theta - \theta^*)| \le c\} d\theta \qquad (3.31) \\
= \;& \frac{2\theta^*}{\pi} F_{2,\nu}(\frac{c^2}{2}) + 2 \int_0^{\frac{\pi}{2}-\theta^*} \frac{1}{2\pi} F_{2,\nu}(\frac{c^2}{2\cos^2\theta}) d\theta \\
& -2 \int_{\frac{\pi}{2}-\theta^*}^{0} \frac{1}{2\pi} F_{2,\nu}(\frac{c^2}{2\cos^2\theta}) d\theta \\
= \;& \frac{2\theta^*}{\pi} F_{2,\nu}(\frac{c^2}{2}) + \frac{2}{\pi} \int_0^{\frac{\pi}{2}-\theta^*} F_{2,\nu}(\frac{c^2}{2\cos^2\theta}) d\theta. \qquad (3.32)
\end{aligned}
$$

3.2.3 Tubular neighborhood method

R in (3.30) can be further written as

$$R = \Big\{\mathbf{t} : \sup_{\mathbf{z} \in \mathcal{Z}} \frac{|\mathbf{z}^T\mathbf{t}|}{\|\mathbf{z}\|\|\mathbf{t}\|} \le \frac{c}{\|\mathbf{t}\|}\Big\}. \qquad (3.33)$$

Then the confidence level is equal to

$$
\begin{aligned}
& \mathrm{P}_{TN}\{\mathbf{t} \in R\} \\
= \;& 1 - \mathrm{P}\Big\{\sup_{\mathbf{z} \in \mathcal{Z}} \frac{|\mathbf{z}^T\mathbf{t}|}{\|\mathbf{z}\|\|\mathbf{t}\|} > \frac{c}{\|\mathbf{t}\|}\Big\} \\
= \;& 1 - \int_{\frac{c^2}{2}}^{\infty} \mathrm{P}\Big\{\sup_{\mathbf{z} \in \mathcal{Z}} \frac{|\mathbf{z}^T\mathbf{t}|}{\|\mathbf{z}\|\|\mathbf{t}\|} > \frac{c}{\sqrt{2w}}\Big\} \cdot dF_{2,\nu}(w). \qquad (3.34)
\end{aligned}
$$

Note that the supreme in (3.34) can not be larger than one. Let $0 < h = c/\sqrt{2w} < 1$ such that $\cos^{-1} h \in (0, \pi/2)$. The set

$$\mathrm{E}(h) = \Big\{\mathbf{t} : \sup_{\mathbf{z} \in \mathcal{Z}} \frac{|\mathbf{z}^T\mathbf{t}|}{\|\mathbf{z}\|\|\mathbf{t}\|} > h\Big\}$$

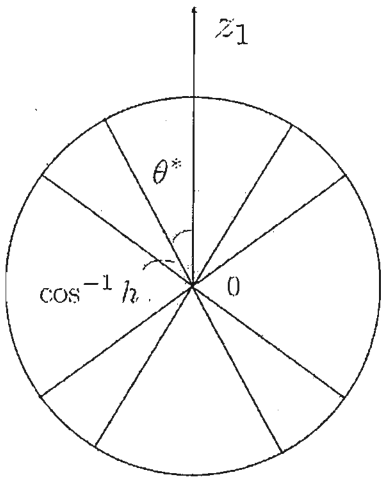

Figure 3.6: For the tubular neighborhood method in two-sided case

consists of two opposite circular cones in $\mathcal{R}^2$ with their common vertex at the origin, symmetrically containing the smaller cones graphically produced by $\mathcal{Z}$ with the angle $\cos^{-1} h$ between the boundaries. So we have $\mathrm{E}(h)$ equal to

$$\begin{array}{ll} \{\mathbf{t} : \theta_{\mathbf{t}} \in [0, \theta^* + \cos^{-1} h] \cup [\pi - \theta^* - \cos^{-1} h] & \text{if} \quad \theta^* + \cos^{-1} h < \frac{\pi}{2}, \\ \quad \cup[\pi + \theta^* + \cos^{-1} h] \cup [2\pi - \theta^* - \cos^{-1} h]\} & \\ \quad\quad \{\mathbf{t} : \theta_{\mathbf{t}} \in [0, 2\pi]\} & \text{if} \quad \theta^* + \cos^{-1} h \geq \frac{\pi}{2}. \end{array}$$

Note that

$$\theta^* + \cos^{-1}(\frac{c}{\sqrt{2w}}) < \frac{\pi}{2} \Longleftrightarrow w < \frac{c^2}{2b^2},$$

where $b = \sqrt{1 - q^2}$. Since $\theta_{\mathbf{t}}$ was proven to be uniformly distributed in Section 3.1, we have, for $c^2/2 \leq w < c^2/(2b^2)$,

$$\begin{aligned} & \mathrm{P}\{\mathbf{t} \in \mathrm{E}(h)\} \\ = \; & \mathrm{P}\{\theta_{\mathbf{t}} \in [0, \theta^* + \cos^{-1}(\frac{c}{\sqrt{2w}})] \cup [\pi - \theta^* - \cos^{-1}(\frac{c}{\sqrt{2w}})] \\ & \quad \cup[\pi + \theta^* + \cos^{-1}(\frac{c}{\sqrt{2w}})] \cup [2\pi - \theta^* - \cos^{-1}(\frac{c}{\sqrt{2w}})]\} \\ = \; & 4 \int_0^{\theta^* + \cos^{-1}(\frac{c}{\sqrt{2w}})} \frac{1}{2\pi} d\theta, \end{aligned} \tag{3.35}$$

and, for $w \geq c^2/(2b^2)$,

$$\mathrm{P}\{\mathrm{t} \in \mathrm{E}(h)\} = \mathrm{P}\{\theta_{\mathrm{t}} \in [0, 2\pi]\} = 1. \tag{3.36}$$

In connection with (3.34), the confidence level is equal to

$$1 - \int_{\frac{c^2}{2}}^{\frac{c^2}{2b^2}} \int_0^{\theta^* + \cos^{-1}(\frac{c}{\sqrt{2w}})} \frac{4}{2\pi} d\theta \cdot dF_{2,\nu}(w) - \int_{\frac{c^2}{2b^2}}^{\infty} 1 \cdot dF_{2,\nu}(w). \tag{3.37}$$

3.2.4 Equivalence of the formulae

We are also interested in finding the equivalence of the formulae (3.29), (3.32) and (3.37) corresponding to the three methods respectively. Obviously (3.29) and (3.32) are the same. So our attention moves to show that (3.32) is equivalent to (3.37).

Consider the double integral in (3.37). It is further equal to

$$\begin{aligned} & \int_{\frac{c^2}{2}}^{\frac{c^2}{2b^2}} \frac{2}{\pi} [\theta^* + \cos^{-1}(\frac{c}{\sqrt{2w}})] dF_{2,\nu}(w) \\ = & \frac{2\theta^*}{\pi} [F_{2,\nu}(\frac{c^2}{2b^2}) - F_{2,\nu}(\frac{c^2}{2})] \\ & + \frac{2}{\pi} \int_{\frac{c^2}{2}}^{\frac{c^2}{2b^2}} \int_0^{\cos^{-1}(\frac{c}{\sqrt{2w}})} d\theta \cdot dF_{2,\nu}(w). \end{aligned} \tag{3.38}$$

Note that $\theta^* = \cos^{-1} q$ and $b = \sqrt{1 - q^2}$ imply $\cos^{-1} b = \pi/2 - \theta^*$. Thus, the last term on the right-hand side in (3.38) can be further written by changing the order of the integrations as

$$\begin{aligned} & \frac{2}{\pi} \int_0^{\cos^{-1} b} \int_{\frac{c^2}{2\cos^2\theta}}^{\frac{c^2}{2b^2}} dF_{2,\nu}(w) \cdot d\theta \\ = & \frac{2}{\pi} \int_0^{\frac{\pi}{2} - \theta^*} [F_{2,\nu}(\frac{c^2}{2b^2}) - F_{2,\nu}(\frac{c^2}{2\cos^2\theta})] d\theta \\ = & (1 - \frac{2\theta^*}{\pi}) F_{2,\nu}(\frac{c^2}{2b^2}) - \frac{2}{\pi} \int_0^{\frac{\pi}{2} - \theta^*} F_{2,\nu}(\frac{c^2}{2\cos^2\theta}) d\theta. \end{aligned} \tag{3.39}$$

Substitute (3.39) into (3.38), and then we have (3.38) further equal to

$$F_{2,\nu}(\frac{c^2}{2b^2}) - \frac{2\theta^*}{\pi} F_{2,\nu}(\frac{c^2}{2}) - \frac{2}{\pi} \int_0^{\frac{\pi}{2} - \theta^*} F_{2,\nu}(\frac{c^2}{2\cos^2\theta}) d\theta. \tag{3.40}$$

By replacing the double integral in (3.37) by (3.40), in connection with that

$$1 - \int_{\frac{c^2}{2b^2}}^{\infty} 1 \cdot dF_{2,\nu}(w) = F_{2,\nu}(\frac{c^2}{2b^2}),$$

we have (3.37) equal to

$$\frac{2\theta^*}{\pi} F_{2,\nu}(\frac{c^2}{2}) + \frac{2}{\pi} \int_0^{\frac{\pi}{2}-\theta^*} F_{2,\nu}(\frac{c^2}{2\cos^2\theta}) d\theta, \qquad (3.41)$$

which is the same as (3.29) and (3.32). Consequently, the equivalence of the three computational formulae is obtained.

Chapter 4

Exact simultaneous confidence bands for a multiple linear regression over an ellipsoidal region

This chapter continues to discuss the construction of exact one-sided and two-sided hyperbolic-shape simultaneous confidence bands for a multiple linear regression model over an ellipsoid that is centered at the point of the means of the predictor variables using the same methods as shown in last chapter. Also, the equivalence of the computational formulae of the methods is given for both one-sided and two-sided cases at last.

4.1 Exact one-sided confidence bands

Bohrer (1973) presented a method of constructing an exact one-sided confidence band for a multiple linear regression model by evaluating a multivariate t probability.

Consider an one-sided hyperbolic-shape simultaneous confidence band e.g., with upper bound, for a classical normal-error multiple linear regression model

$$\mathbf{x}^T\beta \le \mathbf{x}^T\hat{\beta} + r\hat{\sigma}\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}} \qquad \mathbf{x} \in \mathcal{X}, \tag{4.1}$$

where $\hat{\beta}$ is the maximum likelihood estimator of the p-dimensional parameter vector β, $\hat{\sigma}$ is the usual unbiased estimator of σ which is the standard variance of the independent and identically distributed random errors in the linear regression model, the $n\times p$ design matrix X can be expressed by $X = (\mathbf{1}, X_{(1)})$, where $\mathbf{1}$ is the vector containing n ones, $X_{(1)}$ is the $n \times (p-1)$ matrix containing the observed predictor variables, and r is a non-negative critical value.

Consider the restricted region of the predictor space, $\mathcal{X}$, which has the form given by

$$\mathcal{X} = \{\mathbf{x} : \mathbf{x}^T V\mathbf{u}(\mathbf{u}^T V\mathbf{u})^{-1}\mathbf{u}^T V\mathbf{x} \ge c^2\mathbf{x}^T V\mathbf{x}\}, \tag{4.2}$$

where $V = (X^TX)^{-1}$, $\mathbf{u}$ is a p-dimensional vector such that $\mathbf{u}^T V\mathbf{u} = 1$, c is a non-negative constant. Define $\bar{x}_{\cdot j} = \frac{1}{n}\sum_{i=1}^n x_{ij}, j = 2, \ldots, p$ and $\bar{\mathbf{x}}_{(1)} = (\bar{x}_{\cdot 2}, \ldots, \bar{x}_{\cdot p})$. Then $\bar{\mathbf{x}}_{(1)}$ is the mean vector of the observed predictor variables. And, we have

$$X^TX = \begin{pmatrix} n & n\bar{\mathbf{x}}_{(1)}^T \\ n\bar{\mathbf{x}}_{(1)} & X_{(1)}^TX_{(1)} \end{pmatrix}.$$

Furthermore, the inverse is given by

$$V = (X^TX)^{-1} = \begin{pmatrix} \frac{1}{n}(1 + \bar{\mathbf{x}}_{(1)}^T S^{-1}\bar{\mathbf{x}}_{(1)}) & -\frac{1}{n}\bar{\mathbf{x}}_{(1)}^T S^{-1} \\ -\frac{1}{n}S^{-1}\bar{\mathbf{x}}_{(1)} & \frac{1}{n}S^{-1} \end{pmatrix},$$

where

$$\begin{aligned} nS &= (X_{(1)}^T - \bar{\mathbf{x}}_{(1)}\mathbf{1}^T)(X_{(1)} - \mathbf{1}\bar{\mathbf{x}}_{(1)}^T) \\ &= X_{(1)}^TX_{(1)} - n\bar{\mathbf{x}}_{(1)}\bar{\mathbf{x}}_{(1)}^T. \end{aligned}$$

Let $\mathbf{u} = \sqrt{n}(1, \bar{\mathbf{x}}_{(1)}^T)^T$ so that $\mathbf{u}^T V\mathbf{u} = 1$ and $\mathbf{u}^T V\mathbf{x} = 1/\sqrt{n}$. In addition,

$$\mathbf{x}^T V\mathbf{x} = 1 + (\mathbf{x}_{(1)} - \bar{\mathbf{x}}_{(1)})^T S^{-1}(\mathbf{x}_{(1)} - \bar{\mathbf{x}}_{(1)}),$$

where $\mathbf{x}_{(1)}$ is defined such that $\mathbf{x} = (1, \mathbf{x}_{(1)}^T)^T$. Hence, (4.2) can be further written as

$$\begin{aligned} \mathcal{X} &= \{\mathbf{x} : c^2[1 + (\mathbf{x}_{(1)} - \bar{\mathbf{x}}_{(1)})^T S^{-1}(\mathbf{x}_{(1)} - \bar{\mathbf{x}}_{(1)})] \leq 1\} \\ &= \Big\{\mathbf{x} : (\mathbf{x}_{(1)} - \bar{\mathbf{x}}_{(1)})^T S^{-1}(\mathbf{x}_{(1)} - \bar{\mathbf{x}}_{(1)}) \leq \frac{1 - c^2}{c^2}\Big\}, \qquad (4.3) \end{aligned}$$

which is, in fact, an $\bar{\mathbf{x}}_{(1)}$-centered ellipsoid in the predictor space, and whose size can be controlled by $(1 - c^2)/c^2$.

Next, we transform the ellipsoidal region $\mathcal{X}$ to a corresponding region of our interest. For a $\mathbf{u}$, there exists a $p \times (p-1)$ matrix U such that $U^T VU = I_{p-1}$ and $\mathbf{u}^T VU = 0$. So $\mathbf{u}$ and the columns of U, which are linearly independent, form an orthogonal basis of the p-dimensional predictor space. Define

$$\mathbf{z} = \begin{pmatrix} z_1 \\ \mathbf{z}_{(1)} \end{pmatrix} = \begin{pmatrix} \mathbf{u}^T \\ U^T \end{pmatrix} V\mathbf{x}. \qquad (4.4)$$

We have

$$\begin{aligned} \mathbf{x}^T V\mathbf{x} &= \mathbf{x}^T V(\mathbf{u}\ U)(\mathbf{u}\ U)^{-1} V^{-1}[(\mathbf{u}\ U)^T]^{-1}(\mathbf{u}\ U)^T V\mathbf{x} \\ &= \left[\begin{pmatrix} \mathbf{u}^T \\ U^T \end{pmatrix} V\mathbf{x}\right]^T \Big[(\mathbf{u}\ U)^T V(\mathbf{u}\ U)\Big]^{-1} \begin{pmatrix} \mathbf{u}^T \\ U^T \end{pmatrix} V\mathbf{x} \\ &= \begin{pmatrix} z_1 \\ \mathbf{z}_{(1)} \end{pmatrix}^T \begin{pmatrix} z_1 \\ \mathbf{z}_{(1)} \end{pmatrix} = \|\mathbf{z}\|^2. \qquad (4.5) \end{aligned}$$

Since $z_1 = \mathbf{u}^T V\mathbf{x} = 1/\sqrt{n} > 0$ and $\mathbf{u}^T V\mathbf{u} = 1$, therefore, the region $\mathcal{X}$ given in (4.2) can be transformed to the following region

$$\mathrm{E}(c) = \{\mathbf{z} : z_1 \geq c\|\mathbf{z}\|\}. \qquad (4.6)$$

Consequently, any $\mathbf{x}$ belonging to $\mathcal{X}$ definitely has a corresponding $\mathbf{z}$ belonging to $\mathrm{E}(c)$; conversely any $\mathbf{z} \in \mathrm{E}(c)$ corresponds to an $\mathbf{x} \in \mathcal{X}$ as well.

Now, we consider the one-sided confidence band given in (4.1). It has the coverage probability given by

$$\mathrm{P}\Big\{\sup_{\mathbf{x} \in \mathcal{X}} \frac{\mathbf{x}^T(\beta - \hat{\beta})}{\hat{\sigma}\sqrt{\mathbf{x}^T(X^T X)^{-1}\mathbf{x}}} \leq r\Big\},$$

which is equal to

$$\mathrm{P}\Big\{\sup_{\mathbf{x}\in\mathcal{X}} \frac{\mathbf{x}^T V(\mathbf{u}\ U)(\mathbf{u}\ U)^{-1}V^{-1}(\beta-\hat{\beta})}{\hat{\sigma}(\mathbf{x}^T V\mathbf{x})^{1/2}} \le r\Big\}.$$

By recalling the definition of $\mathbf{z}$ in (4.4) and the derivation in (4.5), in connection with the fact that an $\mathbf{x} \in \mathcal{X}$ one-to-one corresponds to a $\mathbf{z} \in \mathrm{E}(c)$, then above probability is equivalent to

$$\mathrm{P}\Big\{\sup_{\mathbf{z}\in\mathrm{E}(c)} \frac{\mathbf{z}^T\mathbf{t}}{||\mathbf{z}||} \le r\Big\}, \tag{4.7}$$

where $\mathbf{t} = \mathrm{N}/\hat{\sigma}$ with $\mathrm{N} = (\mathbf{u}\ U)^{-1}V^{-1}(\beta-\hat{\beta}) = (\mathbf{u}\ U)^T(\beta-\hat{\beta})$. Note that N has the $N_p(0,\sigma^2 I)$ distribution, $\hat{\sigma}$ has the $\sigma\sqrt{\chi^2_\nu/\nu}$ distribution, and they are independent by studying the least squares theory. Thus, $\mathbf{t}$ has a so-called multivariate t distribution with $\nu = n-p$ degrees of freedom. This is the starting point of the three methods given in the rest of this section.

4.1.1 Method of Bohrer (1973)

Define a p-dimensional vector $\mathbf{v}$ in terms of the polar coordinates $R_{\mathbf{v}}$ and $\boldsymbol{\theta}_{\mathbf{v}} = (\theta_{\mathbf{v}1},\ldots,\theta_{\mathbf{v},p-1})$ by

$$\begin{cases} v_1 = R_{\mathbf{v}}\cos\theta_{\mathbf{v}1}, \\ v_2 = R_{\mathbf{v}}\sin\theta_{\mathbf{v}1}\cos\theta_{\mathbf{v}2}, \\ v_3 = R_{\mathbf{v}}\sin\theta_{\mathbf{v}1}\sin\theta_{\mathbf{v}2}\cos\theta_{\mathbf{v}3}, \\ \cdots \quad \cdots \\ v_{p-1} = R_{\mathbf{v}}\sin\theta_{\mathbf{v}1}\sin\theta_{\mathbf{v}2}\cdots\sin\theta_{\mathbf{v},p-2}\cos\theta_{\mathbf{v},p-1}, \\ v_p = R_{\mathbf{v}}\sin\theta_{\mathbf{v}1}\sin\theta_{\mathbf{v}2}\cdots\sin\theta_{\mathbf{v},p-2}\sin\theta_{\mathbf{v},p-1}, \end{cases}$$

where

$$\begin{cases} 0 \le \theta_{\mathbf{v}1} \le \pi, \\ 0 \le \theta_{\mathbf{v}2} \le \pi, \\ \cdots \quad \cdots \\ 0 \le \theta_{\mathbf{v},p-2} \le \pi, \\ 0 \le \theta_{\mathbf{v},p-1} \le 2\pi, \\ R_{\mathbf{v}} \ge 0. \end{cases}$$

Furthermore, the Jacobian of the transformation from $\mathbf{v}$ to the polar coordinates is

$$|J| = R_{\mathbf{v}}^{p-1} \sin^{p-2}\theta_{\mathbf{v}1} \sin^{p-3}\theta_{\mathbf{v}2} \cdots \sin\theta_{\mathbf{v},p-2}.$$

Also, note that

$$\mathbf{z}^T\mathbf{t} = R_{\mathbf{z}} R_{\mathbf{t}} f_{p-1}, \tag{4.8}$$

where $f_1 = \cos(\theta_{\mathbf{z},p-1} - \theta_{\mathbf{t},p-1})$ and

$$f_j = \cos\theta_{\mathbf{z},p-j}\cos\theta_{\mathbf{t},p-j} + f_{j-1}\sin\theta_{\mathbf{z},p-j}\sin\theta_{\mathbf{t},p-j}$$

for $j = 2, \ldots, p-1$.

The probability in (4.7) can be alternatively written as $\mathrm{P}\{\mathbf{t} \in A_r(c)\}$, where

$$A_r = A_r(c) = \{\mathbf{t} : \mathbf{z}^T\mathbf{t} \le r\|\mathbf{z}\|, \text{ all } \mathbf{z} \text{ in } \mathrm{E}(c)\} \tag{4.9}$$

with $\mathrm{E}(c)$ defined in (4.6) which is a spherical cone with the vertex at the origin. $A_r(c)$ and $\mathrm{E}(c)$ are graphically shown in Figure 4.1.

Lemma 4.1.1 *Define $\theta^* = \cos^{-1} c$, then A_r is partitioned by the following three disjoint sets:*

$$\begin{aligned} T_1 &= \{\mathbf{t} : 0 \le \theta_{\mathbf{t}1} \le \theta^*, R_{\mathbf{t}} \le r\}, \\ T_2 &= \{\mathbf{t} : \theta_{\mathbf{t}1} - \theta^* \in (0, \frac{\pi}{2}], R_{\mathbf{t}}\cos(\theta_{\mathbf{t}1} - \theta^*) \le r\}, \\ T_3 &= \{\mathbf{t} : \theta^* + \frac{\pi}{2} < \theta_{\mathbf{t}1} \le \pi\}. \end{aligned} \tag{4.10}$$

Proof. When $\mathbf{t} \in T_1$, then $t_1 = R_{\mathbf{t}}\cos\theta_{\mathbf{t}1} \ge R_{\mathbf{t}}\cos\theta^* = c\|\mathbf{t}\|$ so that $\mathbf{t} \in \mathrm{E}(c)$, which in connection with the fact that $R_{\mathbf{t}}^T R_{\mathbf{t}} \le r\|\mathbf{t}\|$ implies $\mathbf{t} \in A_r(c)$. When $\mathbf{t} \in T_2$, since $\mathbf{z} \in \mathrm{E}(c)$, then $0 \le \theta_{\mathbf{z}1} \le \theta^*$ so that $R_{\mathbf{z}}R_{\mathbf{t}}\cos(\theta_{\mathbf{t}1} - \theta_{\mathbf{z}1}) \le R_{\mathbf{z}}R_{\mathbf{t}}\cos(\theta_{\mathbf{t}1} - \theta^*) \le r\|\mathbf{z}\|$, which implies that $\mathbf{t} \in A_r(c)$. When $\mathbf{t} \in T_3$, obviously $\cos(\theta_{\mathbf{t}1} - \theta_{\mathbf{z}1}) < 0$ which implies $\mathbf{t} \in A_r(c)$. Hence $\cup_{j=1}^3 T_j \subset A_r$.

Conversely, when $\mathbf{t} \in A_r \cap \{\mathbf{t} : 0 \le \theta_{\mathbf{t}1} \le \theta^*\}$, then $t_1 = R_{\mathbf{t}}\cos\theta_{\mathbf{t}1} \ge c\|\mathbf{t}\|$ so that $\mathbf{t} \in \mathrm{E}(c)$ and hence we obtain $R_{\mathbf{t}} \le r$ from $\mathbf{t}^T\mathbf{t} \le r\|\mathbf{t}\|$. Therefore, $\mathbf{t} \in T_1$. When $\mathbf{t} \in A_r \cap \{\mathbf{t} : \theta_{\mathbf{t}1} - \theta^* \in (0, \pi/2)\}$, since $\mathbf{z} \in \mathrm{E}(c)$, we have

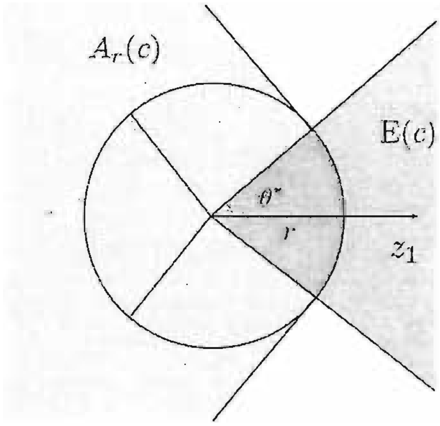

Figure 4.1: For the method of Bohrer (1973)

$R_{\mathbf{t}}\cos(\theta_{\mathbf{t}1}-\theta^*)\le r$. So $\mathbf{t}\in T_2$. And finally $\mathbf{t}\in A_r\cap\{\mathbf{t}:\theta^*+\pi/2<\theta_{\mathbf{t}1}\le\pi\}$ obviously implies $\mathbf{t}\in T_3$. Hence $A_r\subset\cup_{j=1}^3 T_j$. Overall, A_r is composed of $T_j, j=1,2,3$. #

Applying Lemma 4.1.1, we have

$$\mathrm{P}\{\mathbf{t}\in A_r\}=\sum_{j=1}^{3}\mathrm{P}\{\mathbf{t}\in T_j\}. \tag{4.11}$$

Recall that $\mathbf{t}=\mathrm{N}/\hat{\sigma}$, where $\mathrm{N}\sim N_p(0,\sigma^2 I)$, and $\hat{\sigma}\sim\sigma\sqrt{\chi^2_\nu/\nu}$. Moreover, N is independent of $\hat{\sigma}$. Thus $\|\mathbf{t}\|^2/p=(\|\mathrm{N}/\sigma\|^2/p)/(\hat{\sigma}^2/\sigma^2)\sim F_{p,\nu}$, where $F_{p,\nu}$ is the F distribution with p and ν degrees of freedom. Also $\mathbf{t}$ can be expressed in terms of the polar coordinates $R_{\mathbf{t}}$ and $\boldsymbol{\theta}_{\mathbf{t}}=(\theta_{\mathbf{t}1},\ldots,\theta_{\mathbf{t},p-1})$, and N can be expressed in terms of R_{N} and $\boldsymbol{\theta}_{\mathrm{N}}=(\theta_{\mathrm{N}1},\ldots,\theta_{\mathrm{N},p-1})$. Note that $\theta_{\mathbf{t}1},\ldots,\theta_{\mathbf{t},p-1}$ and $\theta_{\mathrm{N}1},\ldots,\theta_{\mathrm{N},p-1}$ denote the same $p-1$ angles because $\mathrm{N}/\hat{\sigma}$ does not change the location of N. One can easily find the joint density function of R_{N} and $\boldsymbol{\theta}_{\mathrm{N}}$ via the transformation of random variables in connection with the fact that N has a p-variate standard normal distribution. By finding the individual marginal density functions, we find that the joint density is equal to the product of the individual marginal densities which implies R_{N} is independent of $\theta_{\mathrm{N}j}, j=1,\ldots,p-1$. Thus, the independence between $R_{\rm t}$ and $\theta_{{\rm t}j}, j = 1, \ldots, p-1$ can be obtained. In particular, $\theta_{\rm t1}$ has the marginal density function

$$f_1(\theta_{\rm t1}) = k_1 \sin^{p-2} \theta_{\rm t1}, \tag{4.12}$$

where k_1 is normalizing constant such that $\int_0^\pi k_1 \sin^{p-2} \theta d\theta = 1$.

Based on the analysis above, we have

$$\begin{aligned}
{\rm P}\{{\rm t} \in T_1\} &= {\rm P}\{0 \leq \theta_{\rm t1} \leq \theta^*, R_{\rm t} \leq r\} \\
&= \int_0^{\theta^*} k_1 \sin^{p-2} \theta d\theta \cdot {\rm P}\{R_{\rm t} \leq r\} \\
&= \int_0^{\theta^*} k_1 \sin^{p-2} \theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}), \qquad (4.13) \\
{\rm P}\{{\rm t} \in T_3\} &= {\rm P}\{\theta^* + \frac{\pi}{2} < \theta_{\rm t1} \leq \pi\} \\
&= \int_{\theta^*+\frac{\pi}{2}}^{\pi} k_1 \sin^{p-2} \theta d\theta \\
&= \int_0^{\frac{\pi}{2}-\theta^*} k_1 \sin^{p-2} \theta d\theta, \qquad (4.14) \\
{\rm P}\{{\rm t} \in T_2\} &= {\rm P}\{0 < \theta_{\rm t1} - \theta^* \leq \frac{\pi}{2}, R_{\rm t} \cos(\theta_{\rm t1} - \theta^*) \leq r\} \\
&= \int_{\theta^*}^{\theta^*+\frac{\pi}{2}} k_1 \sin^{p-2} \theta \cdot {\rm P}\{R_{\rm t} \cos(\theta - \theta^*) \leq r\} d\theta \\
&= \int_0^{\frac{\pi}{2}} k_1 \sin^{p-2}(\theta + \theta^*) \cdot F_{p,\nu}(\frac{r^2}{p \cos^2 \theta}) d\theta, \qquad (4.15)
\end{aligned}$$

where $F_{p,\nu}$ stands for the F cumulative distribution function with p and ν degrees of freedom.

Consequently, by (4.11), the confidence level of the one-sided confidence band based on the method of Bohrer (1973) is given by

$$\begin{aligned}
{\rm P}_B &= \int_0^{\theta^*} k_1 \sin^{p-2} \theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}) + \int_0^{\frac{\pi}{2}-\theta^*} k_1 \sin^{p-2} \theta d\theta + \\
&\quad \int_0^{\frac{\pi}{2}} k_1 \sin^{p-2}(\theta + \theta^*) \cdot F_{p,\nu}(\frac{r^2}{p \cos^2 \theta}) d\theta. \qquad (4.16)
\end{aligned}$$

4.1.2 Algebraical method

Recalling (4.7), the confidence level of the band has the form

$$P\Big\{\sup_{\mathbf{z}\in E(c)} \frac{\mathbf{z}^T\mathbf{t}}{||\mathbf{z}||} \le r\Big\}, \tag{4.17}$$

where r is a critical value, and

$$E(c) = \{\mathbf{z} : z_1 \ge c||\mathbf{z}||\} \tag{4.18}$$

with c non-negative.

Lemma 4.1.2 *Let t_1 be the first element of $\mathbf{t}$, $\mathbf{t}_{(1)}$ be the $(p-1)$-dimensional vector containing the rest elements of $\mathbf{t}$ in order. Then we have*

$$\sup_{\mathbf{z}\in E(c)} \frac{\mathbf{z}^T\mathbf{t}}{||\mathbf{z}||} = \begin{cases} ||\mathbf{t}|| & \textit{if} \quad \mathbf{t} \in E(c), \\ \frac{t_1+\frac{1}{q}||\mathbf{t}_{(1)}||}{\sqrt{1+\frac{1}{q^2}}} & \textit{if} \quad \mathbf{t} \notin E(c), \end{cases}$$

where $q = \sqrt{c^2/(1-c^2)}$.

Proof. Note that it is obvious when $\mathbf{t} \in E(c)$. So our attention focuses on the case when $\mathbf{t} \notin E(c)$. Define $\mathbf{z} = (z_1, \mathbf{z}_{(1)}^T)^T$ such that

$$\begin{aligned} E(c) &= \{\mathbf{z} : z_1 > 0, z_1^2 \ge c^2 z_1^2 + c^2||\mathbf{z}_{(1)}||^2\} \\ &= \{\mathbf{z} : z_1 \ge q||\mathbf{z}_{(1)}||\}, \end{aligned} \tag{4.19}$$

where q is defined in Lemma 4.1.2.

Consider $\mathbf{t} \notin E(c)$ which leads that $t_1 < q||\mathbf{t}_{(1)}||$ from (4.19). For $t_1 \ne 0$, define

$$\mathbf{z}^* = \begin{pmatrix} 1 \\ \frac{\mathbf{t}_{(1)}/|t_1|}{q||\mathbf{t}_{(1)}/|t_1|||} \end{pmatrix},$$

then $z_1^* = 1$ and $q||\mathbf{z}_{(1)}^*|| = 1$ so that $\mathbf{z}^* \in E(c)$. So generally we consider $\mathbf{z} \in E(c)$ has the similar form with $\mathbf{z}^*$ that $\mathbf{z} = (1, \mathbf{z}_{(1)}^T)^T$ and $1 \ge q||\mathbf{z}_{(1)}||$. Therefore, we have

$$\begin{aligned} \frac{\mathbf{z}^T\mathbf{t}}{||\mathbf{z}||} &= \frac{t_1 + \mathbf{z}_{(1)}^T\mathbf{t}_{(1)}}{\sqrt{1+||\mathbf{z}_{(1)}||^2}} \\ &\le \frac{t_1 + ||\mathbf{z}_{(1)}||||\mathbf{t}_{(1)}||}{\sqrt{1+||\mathbf{z}_{(1)}||^2}} \\ &= \phi t_1 + \sqrt{1-\phi^2}||\mathbf{t}_{(1)}|| = f(\phi), \end{aligned} \tag{4.20}$$

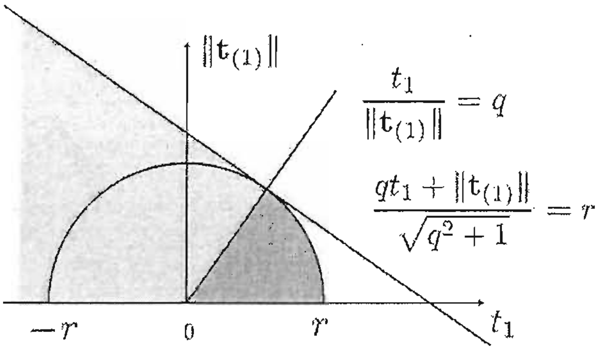

Figure 4.2: For the algebraical method in one-sided case

where

$$\phi = \frac{1}{\sqrt{1 + \|\mathbf{z}_{(1)}\|^2}} \geq \frac{1}{\sqrt{1 + 1/q^2}}.$$

Also, from $t_1 < q\|\mathbf{t}_{(1)}\|$, we have

$$t_1 < \frac{1/\sqrt{1+1/q^2}}{\sqrt{1 - 1/(1+1/q^2)}}\|\mathbf{t}_{(1)}\| \leq \frac{\phi}{\sqrt{1-\phi^2}}\|\mathbf{t}_{(1)}\|. \tag{4.21}$$

So it is clear that $f(\phi)$ is monotonously decreasing because, by (4.20) and (4.21), we have

$$f'(\phi) = t_1 - \frac{1}{2}(1-\phi^2)^{-\frac{1}{2}} \cdot 2\phi\|\mathbf{t}_{(1)}\| \leq 0.$$

Consequently,

$$f(\phi) \leq f(\frac{1}{\sqrt{1+1/q^2}}) = \frac{t_1 + (1/q)\|\mathbf{t}_{(1)}\|}{\sqrt{1+1/q^2}}. \; \# \tag{4.22}$$

The probability in (4.17), therefore, can be evaluated straightforward by applying Lemma 4.1.2. We have

$$\begin{aligned}
& \mathrm{P}\{\sup_{\mathbf{z}\in \mathrm{E}(c)} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|} \le r\} \\
= \; & \mathrm{P}\{\mathbf{t}\in \mathrm{E}(c), \|\mathbf{t}\| \le r\} + \\
& \mathrm{P}\Big\{\mathbf{t}\notin \mathrm{E}(c), \frac{qt_1 + \|\mathbf{t}_{(1)}\|}{\sqrt{q^2+1}} \le r\Big\} \\
= \; & \mathrm{P}\{t_1 \ge q\|\mathbf{t}_{(1)}\|, \|t_1\|^2 + \|\mathbf{t}_{(1)}\|^2 \le r^2\} + \\
& \mathrm{P}\Big\{t_1 < q\|\mathbf{t}_{(1)}\|, \frac{qt_1 + \|\mathbf{t}_{(1)}\|}{\sqrt{q^2+1}} \le r\Big\} \\
= \; & \mathrm{P}\{R_{\mathbf{t}}^2 \le r^2\} + \mathrm{P}\Big\{r^2 < R_{\mathbf{t}}^2 < \infty, \\
& w = \frac{t_1}{\|\mathbf{t}_{(1)}\|} \in \Big(-\infty, \frac{ar - b\sqrt{R_{\mathbf{t}}^2 - r^2}}{br + a\sqrt{R_{\mathbf{t}}^2 - r^2}}\Big)\Big\}, \qquad (4.23)
\end{aligned}$$

where $R_{\mathbf{t}}^2 = t_1^2 + \|\mathbf{t}_{(1)}\|^2$, $a = q/\sqrt{q^2+1}$, $b = 1/\sqrt{q^2+1}$, and the upper bound of w is obtained by solving the equation set formed by $\|t_1\|^2 + \|\mathbf{t}_{(1)}\|^2 = r^2$ and $(qt_1 + \|\mathbf{t}_{(1)}\|)/\sqrt{q^2+1} = r$. The accomplishment of the last equality in (4.23) is graphically because the total area of the light shadowed region and the dark shadowed region is equal to the total area of the half circle and the rest light shadowed part in the left top corner. The first probability on the right-hand side of the last equality in (4.23) equals $F_{p,\nu}(r^2/p)$ which is the F cumulative distribution function with p and ν degrees of freedom, and the second probability is further equal to

$$\int_{\frac{r^2}{p}}^{\infty} g(w)dF_{p,\nu}(w), \qquad (4.24)$$

where

$$g(w) = \mathrm{P}\Big\{\frac{t_1}{\|\mathbf{t}_{(1)}\|} \le \frac{ar - b\sqrt{pw - r^2}}{br + a\sqrt{pw - r^2}}\Big\}.$$

To evaluate $g(w)$, note that

$$ar - b\sqrt{pw - r^2} \le 0 \Longleftrightarrow w \ge r^2/pb^2.$$

For $w \in (r^2/p, r^2/pb^2)$, $(ar - b\sqrt{pw - r^2})/(br + a\sqrt{pw - r^2}) \geq 0$, then

$$\begin{aligned}
g(w) &= \mathrm{P}\{t_1 \leq 0\} + \mathrm{P}\Big\{0 \leq \frac{t_1}{\|\mathbf{t}_{(1)}\|} \leq \frac{ar - b\sqrt{pw - r^2}}{br + a\sqrt{pw - r^2}}\Big\} \\
&= \frac{1}{2} + \frac{1}{2}\mathrm{P}\Big\{\frac{t_1^2}{\|\mathbf{t}_{(1)}\|^2/(p-1)} \leq (p-1)\Big(\frac{ar - b\sqrt{pw - r^2}}{br + a\sqrt{pw - r^2}}\Big)^2\Big\} \\
&= \frac{1}{2} + \frac{1}{2}F_{1,p-1}\Big\{(p-1)\Big(\frac{ar - b\sqrt{pw - r^2}}{br + a\sqrt{pw - r^2}}\Big)^2\Big\}; \qquad (4.25)
\end{aligned}$$

for $w \in (r^2/pb^2, \infty)$, $(ar - b\sqrt{pw - r^2})/(br + a\sqrt{pw - r^2}) \leq 0$, then

$$\begin{aligned}
g(w) &= \mathrm{P}\Big\{\frac{t_1^2}{\|\mathbf{t}_{(1)}\|^2} \geq (\frac{ar - b\sqrt{pw - r^2}}{br + a\sqrt{pw - r^2}})^2\Big\} \\
&= \frac{1}{2} - \frac{1}{2}\mathrm{P}\Big\{\frac{t_1^2}{\|\mathbf{t}_{(1)}\|^2/(p-1)} \leq (p-1)\Big(\frac{ar - b\sqrt{pw - r^2}}{br + a\sqrt{pw - r^2}}\Big)^2\Big\} \\
&= \frac{1}{2} - \frac{1}{2}F_{1,p-1}\Big\{(p-1)\Big(\frac{ar - b\sqrt{pw - r^2}}{br + a\sqrt{pw - r^2}}\Big)^2\Big\}, \qquad (4.26)
\end{aligned}$$

where $a = q/\sqrt{q^2+1} = c$, $b = 1/\sqrt{q^2+1} = \sqrt{1-c^2}$.

4.1.3 Tubular neighborhood method

Again from (4.7), the confidence level of the one-sided confidence band can be further written as

$$\begin{aligned}
&\mathrm{P}\Big\{\sup_{\mathbf{z}\in \mathrm{E}(c)} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|} \leq r\Big\} \\
=\ & 1 - \int_0^\infty \mathrm{P}\Big\{\sup_{\mathbf{z}\in \mathrm{E}(c)} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|\|\mathbf{t}\|} > \frac{r}{\sqrt{pw}}\Big\} dF_{p,\nu}(w) \\
=\ & 1 - \int_{\frac{r^2}{p}}^\infty \mathrm{P}\Big\{\sup_{\mathbf{z}\in \mathrm{E}(c)} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|\|\mathbf{t}\|} > \frac{r}{\sqrt{pw}}\Big\} dF_{p,\nu}(w). \qquad (4.27)
\end{aligned}$$

Note that the supreme on the right-hand side of the last equality in (4.27) is no larger than one. Let $0 < h = r/\sqrt{pw} < 1$ such that $\cos^{-1} h \in (0, \pi/2)$, the set

$$\mathrm{E}(h) = \Big\{\mathbf{t} : \sup_{\mathbf{z}\in \mathrm{E}(c)} \frac{\mathbf{z}^T\mathbf{t}}{\|\mathbf{z}\|\|\mathbf{t}\|} > h\Big\} \qquad (4.28)$$

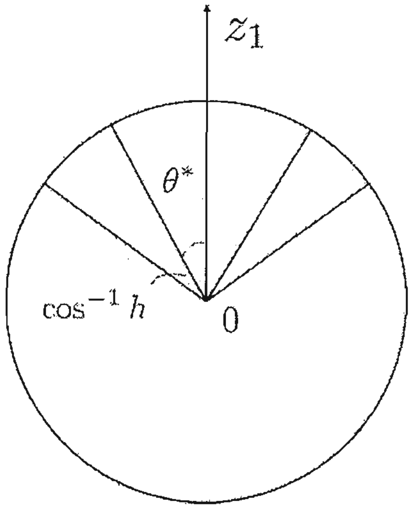

Figure 4.3: For the tubular neighborhood method in one-sided case

graphically is a spherical cone with the vertex at the origin and has the angle $\cos^{-1} h$ between a ray of $\mathrm{E}(h)$ and the nearest ray from $\mathrm{E}(c)$. By the definition of the polar coordinates, $\mathrm{P}\{\mathbf{t} \in \mathrm{E}(h)\}$ is equal to $\mathrm{P}\{0 < \theta_{\mathbf{t},1} \leq \theta^* + \cos^{-1} h\}$, and $\|\mathbf{t}\| = R_{\mathbf{t}}$ is independent of $\theta_{\mathbf{t}1}$. Therefore, (4.27) is further equal to

$$\begin{aligned} & 1 - \int_{\frac{r^2}{p}}^{\infty} \mathrm{P}\{0 < \theta_{\mathbf{t},1} \leq \theta^* + \cos^{-1}(\frac{r}{\sqrt{pw}})\} dF_{p,\nu}(w) \\ = \; & 1 - \int_{\frac{r^2}{p}}^{\infty} \int_0^{\theta^* + \cos^{-1}(\frac{r}{\sqrt{pw}})} k_1 \sin^{p-2} \theta d\theta \cdot dF_{p,\nu}(w), \end{aligned} \tag{4.29}$$

where

$$k_1 = \frac{1}{\int_0^\pi \sin^{p-2} \theta d\theta} \tag{4.30}$$

is the normalizing constant.

4.1.4 Equivalence of the formulae

It is of interest to show the equivalence of (4.16), (4.29) and that based on the algebraical method. First, we come to show (4.29) is equivalent to (4.16).

Further write (4.29) as

$$
\begin{aligned}
& 1 - \int_{\frac{r^2}{p}}^{\infty} \Big(\int_0^{\theta^*} + \int_{\theta^*}^{\theta^* + \cos^{-1}(\frac{r}{\sqrt{pw}})} \Big) k_1 \sin^{p-2}\theta d\theta \cdot dF_{p,\nu}(w) \\
= \; & 1 - \int_{\frac{r^2}{p}}^{\infty} dF_{p,\nu}(w) \int_0^{\theta^*} k_1 \sin^{p-2}\theta d\theta \\
& - \int_{\frac{r^2}{p}}^{\infty} \int_0^{\cos^{-1}(\frac{r}{\sqrt{pw}})} k_1 \sin^{p-2}(\theta + \theta^*) d\theta \cdot dF_{p,\nu}(w) \\
= \; & \Big(1 - \int_0^{\theta^*} k_1 \sin^{p-2}\theta d\theta \Big) + \int_0^{\theta^*} k_1 \sin^{p-2}\theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}) \\
& - \int_0^{\frac{\pi}{2}} \int_{\frac{r^2}{p\cos^2\theta}}^{\infty} k_1 \sin^{p-2}(\theta + \theta^*) dF_{p,\nu}(w) \cdot d\theta \\
= \; & \int_0^{\theta^*} k_1 \sin^{p-2}\theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}) + \int_0^{\frac{\pi}{2}} k_1 \sin^{p-2}(\theta + \theta^*) F_{p,\nu}(\frac{r^2}{p\cos^2\theta}) d\theta \\
& + \Big(1 - \int_0^{\theta^*} k_1 \sin^{p-2}\theta d\theta - \int_0^{\frac{\pi}{2}} k_1 \sin^{p-2}(\theta + \theta^*) d\theta \Big). \qquad (4.31)
\end{aligned}
$$

The terms in the big bracket on the right-hand side of the last equality in (4.31) together as a whole is further equal to

$$
\Big(\int_0^{\pi} - \int_0^{\theta^*} - \int_{\theta^*}^{\theta^* + \frac{\pi}{2}} \Big) k_1 \sin^{p-2}\theta d\theta = \int_0^{\frac{\pi}{2} - \theta^*} k_1 \sin^{p-2}\theta d\theta. \qquad (4.32)
$$

Substituting (4.32) into (4.31) gives the same formula as (4.16).

Next, we turn to find the equivalence between (4.16) and that got from the algebraical method. Recall (4.23), with θ^* defined consistently, the last equality is equal to

$$\begin{aligned}
& \mathrm{P}\{R_{\mathrm{t}}^2 \le r^2, \theta_{\mathrm{t}1} \in [0,\theta^*]\} + \mathrm{P}\{R_{\mathrm{t}}\cos(\theta_{\mathrm{t}1}-\theta^*) \le r, \theta_{\mathrm{t}1} \in (\theta^*,\pi]\} \\
= & \mathrm{P}\{R_{\mathrm{t}}^2 \le r^2, \theta_{\mathrm{t}1} \in [0,\theta^*]\} + \mathrm{P}\{0 \le R_{\mathrm{t}}\cos(\theta_{\mathrm{t}1}-\theta^*) \le r, \theta_{\mathrm{t}1} \in (\theta^*,\pi]\} \\
& +\mathrm{P}\{R_{\mathrm{t}}\cos(\theta_{\mathrm{t}1}-\theta^*) < 0, \theta_{\mathrm{t}1} \in (\theta^*,\pi]\} \\
= & \mathrm{P}\{R_{\mathrm{t}}^2 \le r^2, \theta_{\mathrm{t}1} \in [0,\theta^*]\} + \mathrm{P}\{0 \le R_{\mathrm{t}}\cos(\theta_{\mathrm{t}1}-\theta^*) \le r, \theta_{\mathrm{t}1} \in (\theta^*,\theta^*+\frac{\pi}{2}]\} \\
& +\mathrm{P}\{R_{\mathrm{t}}\cos(\theta_{\mathrm{t}1}-\theta^*) < 0, \theta_{\mathrm{t}1} \in (\theta^*+\frac{\pi}{2},\pi]\} \\
= & \int_0^{\theta^*} k_1 \sin^{p-2}\theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}) + \int_{\theta^*}^{\theta^*+\frac{\pi}{2}} k_1 \sin^{p-2}\theta \cdot F_{p,\nu}(\frac{r^2}{p\cos^2(\theta-\theta^*)})d\theta \\
& + \int_{\theta^*+\frac{\pi}{2}}^{\pi} k_1 \sin^{p-2}\theta d\theta \\
= & \int_0^{\theta^*} k_1 \sin^{p-2}\theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}) + \int_0^{\frac{\pi}{2}} k_1 \sin^{p-2}(\theta+\theta^*) \cdot F_{p,\nu}(\frac{r^2}{p\cos^2\theta})d\theta \\
& + \int_0^{\frac{\pi}{2}-\theta^*} k_1 \sin^{p-2}\theta d\theta.
\end{aligned} \quad (4.33)$$

Obviously, (4.33) is the same as (4.16). So a conclusion can be drawn that the three methods give the same computational formulae.

4.2 Exact two-sided confidence bands

In this section, we consider exact two-sided simultaneous confidence bands for a multiple linear regression over an ellipsoidal region based on the same methods.

4.2.1 Method following the idea of Bohrer (1973)

Recall (4.6) and (4.9). We change $\mathrm{E}(c)$ and $A_r(c)$ slightly to make them correspond to the two-sided case. Therefore, we have for two-sided confidence bands

$$A_r = \{\mathbf{t} : |\mathbf{z}^T\mathbf{t}| \le r\|\mathbf{z}\|, \text{ all } \mathbf{z} \text{ in } \mathrm{E}(c)\}, \quad (4.34)$$

where

$$\mathrm{E}(c) = \{\mathbf{z} : |z_1| \ge c\|\mathbf{z}\|\}.$$

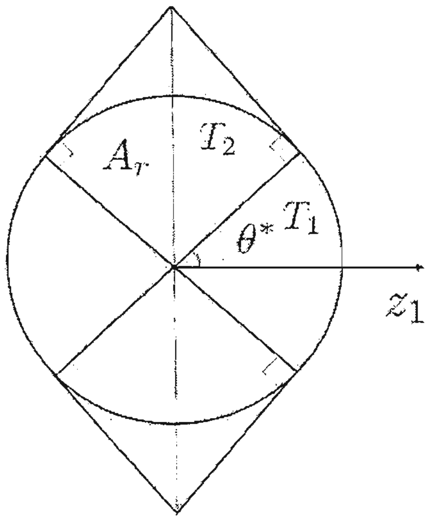

Figure 4.4: For the method following Bohrer (1973) in two-sided case

Recall the definitions of the polar coordinates $(R_{\mathbf{t}}, \theta_{\mathbf{t}1}, \ldots, \theta_{\mathbf{t},p-1})$, where $\theta_{\mathbf{t}j} \in [0, \pi], j = 1, \ldots, p-2$ and $\theta_{\mathbf{t},p-1} \in [0, 2\pi]$. Note that when we consider $\theta_{\mathbf{t}1}$ moving throughout $[0, \pi]$, A_r actually looks like the full region rather than just the upper half due to the effects of other angles $\theta_{\mathbf{t}j}, j = 2, \ldots, p-1$. Also, note that A_r has a graphically symmetric shape for the two-sided case as shown in Figure 4.4. So we only need to consider the region produced by $\theta_{\mathbf{t}1}$ moving throughout $[0, \pi/2]$.

Define

$$\begin{aligned}
\theta^* &= \cos^{-1} c, \\
T_1 &= \{\mathbf{t} : 0 \le \theta_{\mathbf{t},1} \le \theta^*, R_{\mathbf{t}} \le r\}, \\
T_2 &= \{\mathbf{t} : \theta^* \le \theta_{\mathbf{t},1} \le \frac{\pi}{2}, R_{\mathbf{t}} \cos(\theta_{\mathbf{t},1} - \theta^*) \le r\}. \qquad (4.35)
\end{aligned}$$

We have the confidence level of the two-sided band simply equal to

$$\mathrm{P}\{\mathbf{t} \in A_r\} = 2\Big(\mathrm{P}\{\mathbf{t} \in T_1\} + \mathrm{P}\{\mathbf{t} \in T_2\}\Big). \qquad (4.36)$$

Recall that $\theta_{\mathbf{t}1}$ has the density function given by

$$f(\theta) = k_1 \sin^{p-2} \theta \qquad (4.37)$$

with k_1 being the normalizing constant, R_t^2/p has the $F_{p,\nu}$ distribution and is independent of θ_{t1}. Thus, we obtain

$$\begin{aligned}
P\{t \in T_1\} &= P\{0 \le \theta_{t1} \le \theta^*, R_t \le r\} \\
&= \int_0^{\theta^*} k_1 \sin^{p-2}\theta d\theta \cdot P\{R_t \le r\} \\
&= \int_0^{\theta^*} k_1 \sin^{p-2}\theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}), \qquad (4.38) \\
P\{t \in T_2\} &= P\{\theta^* \le \theta_{t1} \le \frac{\pi}{2}, R_t \cos(\theta_{t1} - \theta^*) \le r\} \\
&= \int_{\theta^*}^{\frac{\pi}{2}} k_1 \sin^{p-2}\theta d\theta \cdot P\{R_t \cos(\theta_{t1} - \theta^*) \le r\} \\
&= \int_0^{\frac{\pi}{2}-\theta^*} k_1 \sin^{p-2}(\theta + \theta^*) d\theta \cdot P\{R_t \le \frac{r}{\cos\theta}\} \\
&= \int_0^{\frac{\pi}{2}-\theta^*} k_1 \sin^{p-2}(\theta + \theta^*) \cdot F_{p,\nu}(\frac{r^2/p}{\cos^2\theta}) d\theta. \qquad (4.39)
\end{aligned}$$

Overall, the two-sided simultaneous confidence band can be constructed with the confidence level given by

$$\int_0^{\theta^*} 2k_1 \sin^{p-2}\theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}) + \int_0^{\frac{\pi}{2}-\theta^*} 2k_1 \sin^{p-2}(\theta+\theta^*) \cdot F_{p,\nu}(\frac{r^2/p}{\cos^2\theta}) d\theta. \quad (4.40)$$

4.2.2 Algebraical method

Casella and Strawderman (1980) considered the construction of a two-sided hyperbolic-shape confidence band over an ellipsoidal region $\mathcal{X}$. And the structure of this $\mathcal{X}$ can be transformed as

$$E(q) = \{\mathbf{z} : \sum_{i=1}^{m} z_i^2 \ge q^2 \sum_{i=m+1}^{p} z_i^2\}.$$

Specially when $m = 1$, $E(q)$ becomes the structure of our interest that

$$E(q) = \{\mathbf{z} : |z_1| \ge \frac{q}{\sqrt{1+q^2}} \|\mathbf{z}\|\},$$

where $q > 0$ is a fixed constant.

From the result of Casella and Strawderman (1980), it is that the two-sided confidence band of hyperbolic shape

$$\mathbf{x}^T\beta \in \mathbf{x}^T\hat{\beta} \pm r\hat{\sigma}\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}} \quad \mathbf{x} \in \mathcal{X}, \tag{4.41}$$

where r is a critical value, has its simultaneous confidence level given by

$$F_{p,\nu}(\frac{r^2}{p}) + \int_{r^2/p}^{r^2/(b^2p)} F_{1,p-1}\Big\{(p-1)\Big(\frac{cr - b\sqrt{pw - r^2}}{br + c\sqrt{pw - r^2}}\Big)^2\Big\} dF_{p,\nu}(w), \tag{4.42}$$

where $c = q/\sqrt{1+q^2}$ and $b = 1/\sqrt{1+q^2}$.

4.2.3 Tubular neighborhood method

Recall (4.7), the confidence level of the two-sided band is alternatively given by

$$\mathrm{P}\Big\{\sup_{\mathbf{z}\in \mathrm{E}(c)} \frac{|\mathbf{z}^T\mathbf{t}|}{\|\mathbf{z}\|\|\mathbf{t}\|} \leq \frac{r}{\|\mathbf{t}\|}\Big\}. \tag{4.43}$$

Note that $\mathbf{t}/\|\mathbf{t}\|$ is independent of $\|\mathbf{t}\|$, the supreme in (4.43) is no larger than one, and

$$r/\sqrt{pw} < 1 \Longleftrightarrow w > r^2/p.$$

Therefore, (4.43) is further equal to

$$\begin{aligned} & 1 - \int_0^\infty \mathrm{P}\Big\{\sup_{\mathbf{y}\in \mathrm{E}(c)} \frac{|\mathbf{y}^T\mathbf{t}|}{\|\mathbf{y}\|\|\mathbf{t}\|} > \frac{r}{\sqrt{pw}}\Big\} dF_{p,\nu}(w) \\ = \; & 1 - \int_{r^2/p}^\infty \mathrm{P}\Big\{\sup_{\mathbf{y}\in \mathrm{E}(c)} \frac{|\mathbf{y}^T\mathbf{t}|}{\|\mathbf{y}\|\|\mathbf{t}\|} > \frac{r}{\sqrt{pw}}\Big\} dF_{p,\nu}(w). \end{aligned} \tag{4.44}$$

Let $0 < h = r/\sqrt{pw} < 1$ such that $\cos^{-1} h \in (0, \pi/2)$. The set

$$\mathrm{E}(h) = \Big\{\mathbf{t} : \sup_{\mathbf{z}\in \mathrm{E}(c)} \frac{|\mathbf{z}^T\mathbf{t}|}{\|\mathbf{z}\|\|\mathbf{t}\|} > h\Big\}$$

consists of two opposite spherical cones in $\mathcal{R}^p$. One cone, C, has its vertex at the origin and its central direction given by z_1-axis, symmetrically and centrally containing one smaller cone with the angle $\cos^{-1} h$ between a ray on its surface and the nearest ray from the smaller cone. The other cone is

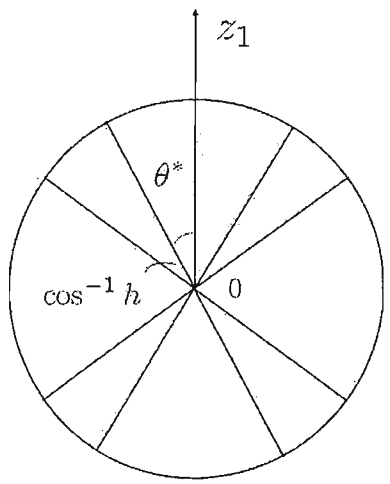

Figure 4.5: For the tubular neighborhood method in two-sided case

simply $-\mathcal{C}$. The two smaller cones are produced by $\mathrm{E}(c)$. Then, in connection with the definition of the polar coordinates, we have $\mathrm{E}(h)$ equal to

$$\begin{array}{ll} \{\mathbf{t} : \theta_{t1} \in [0, \theta^* + \cos^{-1} h] \cup [\pi - \theta^* - \cos^{-1} h, \pi]\} & \text{if} \quad \theta^* + \cos^{-1} \mathrm{h} < \frac{\pi}{2}, \\ \{\mathbf{t} : \theta_{t1} \in [0, \pi]\} & \text{if} \quad \theta^* + \cos^{-1} \mathrm{h} \geq \frac{\pi}{2}. \end{array}$$

Note that

$$\theta^* + \cos^{-1}(r/\sqrt{pw}) < \frac{\pi}{2} \Longleftrightarrow w < \frac{r^2}{b^2 p},$$

and the density function of $\theta_{t,1}$ is given by (4.37), we therefore have, for $r^2/p \leq w < r^2/(b^2 p)$,

$$\begin{aligned} & \mathrm{P}\{\mathbf{t} \in \mathrm{E}(h)\} \\ = \; & \mathrm{P}\{\theta_{t,1} \in [0, \theta^* + \cos^{-1}(r/\sqrt{pw})] \cup [\pi - \theta^* - \cos^{-1}(r/\sqrt{pw}), \pi]\} \\ = \; & \int_0^{\theta^* + \cos^{-1}(r/\sqrt{pw})} 2k_1 \sin^{p-2} \theta d\theta; \end{aligned}$$

and, for $w \geq r^2/(b^2 p)$,

$$\mathrm{P}\{\mathbf{t} \in \mathrm{E}(h)\} = \mathrm{P}\{\theta_{t,1} \in [0, \pi]\} = 1.$$

Consequently, the confidence level (4.44) is equal to

$$
\begin{aligned}
1 &- \int_{r^2/p}^{r^2/(b^2p)} \int_0^{\theta^*+\cos^{-1}(r/\sqrt{pw})} 2k_1 \sin^{p-2}\theta d\theta dF_{p,\nu}(w) \\
&- \int_{r^2/(b^2p)}^{\infty} 1 dF_{p,\nu}(w). \qquad (4.45)
\end{aligned}
$$

4.2.4 Equivalence of the formulae

By changing the order of integrations, the double integral in (4.45) simplifies as

$$
\begin{aligned}
& \int_0^{\theta^*} 2k_1 \sin^{p-1}\theta d\theta \{F_{p,\nu}(\frac{r^2}{b^2p}) - F_{p,\nu}(\frac{r^2}{p})\} \\
+ & \int_{\theta^*}^{\frac{\pi}{2}} 2k_1 \sin^{p-1}\theta \Big\{ F_{p,\nu}(\frac{r^2}{b^2p}) - F_{p,\nu}\Big(\frac{r^2}{p\cos^2(\theta-\theta^*)}\Big)\Big\} d\theta. \qquad (4.46)
\end{aligned}
$$

Substitute (4.46) into (4.45) and note that

$$
\int_0^{\frac{\pi}{2}} 2k_1 \sin^{p-1}\theta d\theta \cdot F_{p,\nu}(\frac{r^2}{b^2p}) = \int_0^{r^2/(b^2p)} dF_{p,\nu}(w).
$$

We finally have that the confidence level based on the tubular neighborhood method is equivalent to the expression given in (4.40).

Change (4.23) slightly to make it corresponding to the two-sided case. We therefore have the confidence level of the two-sided band is

$$
\begin{aligned}
& \mathrm{P}\{R_{\mathbf{t}}^2 \le r^2\} + \mathrm{P}\Big\{ r^2 < R_{\mathbf{t}}^2 \le \frac{r^2}{b^2}, w = \frac{t_1}{\|\mathbf{t}_{(1)}\|} \\
& \in \Big(-\frac{ar - b\sqrt{R_{\mathbf{t}}^2 - r^2}}{br + a\sqrt{R_{\mathbf{t}}^2 - r^2}}, \frac{ar - b\sqrt{R_{\mathbf{t}}^2 - r^2}}{br + a\sqrt{R_{\mathbf{t}}^2 - r^2}} \Big) \Big\}, \qquad (4.47)
\end{aligned}
$$

where $a = q/\sqrt{1+q^2}$ and $b = 1/\sqrt{1+q^2}$. The first probability is simply equal to $F_{p,\nu}(r^2/p)$, and the second one is further equal to

$$
\begin{aligned}
& \int_{\frac{r^2}{p}}^{\frac{r^2}{b^2p}} \mathrm{P}\Big\{ \frac{|t_1|}{\|\mathbf{t}_{(1)}\|} \le \frac{ar - b\sqrt{pw - r^2}}{br + a\sqrt{pw - r^2}} \Big\} dF_{p,\nu}(w) \\
= & \int_{\frac{r^2}{p}}^{\frac{r^2}{b^2p}} F_{1,p-1}\Big\{ (p-1)\Big(\frac{ar - b\sqrt{pw - r^2}}{br + a\sqrt{pw - r^2}}\Big)^2 \Big\} dF_{p,\nu}(w). \qquad (4.48)
\end{aligned}
$$

It can be found that (4.47) is equivalent to (4.42). On the other hand, we can write (4.47) according to the location of the vector $\mathbf{t}$. It is that (4.47) can be evaluated in terms of the regions θ_{t1} belongs to, as

$$
\begin{aligned}
& \mathrm{P}\{R_t^2 \leq r^2, \theta_{t1} \in [0, \theta^*] \cup [\pi - \theta^*, \pi]\} \\
& + \mathrm{P}\{R_t \cos(\theta_{t1} - \theta^*) \leq r, \theta_{t1} \in (\theta^*, \pi - \theta^*)\} \\
= & \int_0^{\theta^*} k_1 \sin^{p-2}\theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}) \\
& + \int_{\pi-\theta^*}^{\pi} k_1 \sin^{p-2}\theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}) \\
& + \int_{\theta^*}^{\pi-\theta^*} k_1 \sin^{p-2}\theta \cdot F_{p,\nu}(\frac{r^2}{p\cos^2(\theta - \theta^*)}) d\theta \\
= & \int_0^{\theta^*} 2k_1 \sin^{p-2}\theta d\theta \cdot F_{p,\nu}(\frac{r^2}{p}) \\
& + \int_0^{\frac{\pi}{2}-\theta^*} 2k_1 \sin^{p-2}(\theta + \theta^*) \cdot F_{p,\nu}(\frac{r^2}{p\cos^2\theta}) d\theta, \qquad (4.49)
\end{aligned}
$$

which is the same as (4.40). Consequently, we obtain the equivalence of the three formulae corresponding to the three methods respectively.

Chapter 5

Simultaneous confidence bands for a regression model over a rectangular region and comparisons

In last two chapters, we discussed the construction of exact simultaneous confidence bands with the predictor variables restricted in an ellipsoidal region. In this chapter, we turn to consider the construction of two-sided simultaneous confidence bands over the most popular rectangular region of the predictor space based on several methods, including Naiman (1986)'s conservative method by applying the tube volume theory, the approximate method proposed by Sun and Loader (1994) presenting an approximation to the tube formula to construct confidence bands for a parametric or nonparametric regression function, and the simulation-based method of Liu, Wynn and Hayter (2005) and Liu, Jamshidian, Zhang and Donnelly (2005) to construct confidence bands for polynomial regression and multiple linear regression respectively. Also, comparisons for these methods are given in terms of critical values. All critical values are calculated by running programmes on MATLAB 7 platform. Conclusions are drawn in the end.

5.1 Conservative confidence bands

Naiman (1986) presented a method of constructing conservative hyperbolic-shape simultaneous confidence bands for an one-dimensional curvilinear regression over finite intervals. This method is, by using a geometric inequality, to obtain an upper bound for the volume of a tube with a fixed distance from an arbitrary path which is piecewise differentiable and has a finite length on the surface S^{p-1} of the unit sphere in p-dimensional real space.

Consider the regression model

$$y = \mathrm{f}(x)^T \boldsymbol{\beta} + \varepsilon, \tag{5.1}$$

where y is the response, $\mathrm{f}(x)$ is the p-dimensional vector of known functions of the only predictor variable x, $\boldsymbol{\beta}$ is the p-dimensional vector of unknown regression coefficients, ε is the random error which is normally distributed with mean 0 and unknown variance σ^2. For a special case when $\mathrm{f}(x) = (1, x, x^2, \ldots, x^{p-1})^T$, (5.1) is the usual polynomial regression model of $p-1$ degrees. Let $\mathcal{X} \subset \mathcal{R}$ be a restricted interval containing all possible values of x. Denote $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ the maximum likelihood estimator of $\boldsymbol{\beta}$ and the usual unbiased estimator of σ^2 respectively. Also, assume the design matrix is of full rank so that $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2\Sigma)$ for some known positive definite matrix Σ. And $\nu\hat{\sigma}^2/\sigma^2 \sim \chi^2_\nu$ with ν degrees of freedom. Let P be a $p \times p$ non-singular matrix such that $P^T P = \Sigma$.

A two-sided hyperbolic-shape simultaneous confidence band for the mean regression function $\mathrm{f}(x)^T\boldsymbol{\beta}$ over the restricted predictor space is given by

$$\mathrm{f}(x)^T\boldsymbol{\beta} \in \mathrm{f}(x)^T\hat{\boldsymbol{\beta}} \pm c\hat{\sigma}p(x) \quad \text{all } x \in \mathcal{X}, \tag{5.2}$$

where $p(x) = \{\mathrm{f}(x)^T\Sigma\mathrm{f}(x)\}^{1/2} = \|P\mathrm{f}(x)\|$, and $c \geq 0$ is a critical value. Define $\mathrm{N} = (P^{-1})^T(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})/\sigma$ which in fact has a p-variate standard normal distribution by studying the least squares theory. Then the confidence band

(5.2) has the confidence level given by

$$
\begin{aligned}
& \mathrm{P}\Big\{ \sup_{x\in\mathcal{X}} \frac{|\mathrm{f}(x)^T(\beta - \hat{\beta})|}{p(x)} \le c\hat{\sigma} \Big\} \\
= \ & \mathrm{P}\Big\{ \sup_{x\in\mathcal{X}} \frac{|[P\mathrm{f}(x)]^T(P^T)^{-1}(\beta - \hat{\beta})/\sigma|}{\|P\mathrm{f}(x)\|} \le \frac{c\hat{\sigma}}{\sigma} \Big\} \\
= \ & \mathrm{P}\Big\{ \sup_{x\in\mathcal{X}} \frac{|[P\mathrm{f}(x)]^T\mathbf{N}|}{\|P\mathrm{f}(x)\|\|\mathbf{N}\|} \le \frac{c\hat{\sigma}/\sigma}{\|\mathbf{N}\|} \Big\} \\
= \ & 1 - \mathrm{P}\Big\{ \sup_{x\in\mathcal{X}} \frac{|[P\mathrm{f}(x)]^T\mathbf{N}|}{\|P\mathrm{f}(x)\|\|\mathbf{N}\|} > \frac{c\hat{\sigma}/\sigma}{\|\mathbf{N}\|} \Big\}. \qquad (5.3)
\end{aligned}
$$

Define

$$\gamma(x) = \frac{P\mathrm{f}(x)}{\|P\mathrm{f}(x)\|} \quad \text{for } x \in \mathcal{X}, \qquad (5.4)$$

which is a path in S^{p-1}, the surface of the unit sphere centered at the origin in $\mathcal{R}^p$. And assume the length of the path given by $\Lambda(\gamma) = \int_{\mathcal{X}} \|\gamma'(x)\|dx$ is finite. Also, define the random vector $\mathbf{U} = \mathbf{N}/\|\mathbf{N}\|$. Consider $\mathbf{N}$ in terms of the polar coordinates $R_{\mathbf{N}} = \|\mathbf{N}\|$ and $\boldsymbol{\theta}_{\mathbf{N}}$. Then $\mathbf{U}$ only depends on $\boldsymbol{\theta}_{\mathbf{N}}$. Since $\mathbf{N}$ has the p-variate standard normal distribution, one may directly find the joint density function of $\|\mathbf{N}\|$ and $\boldsymbol{\theta}_{\mathbf{N}}$. Furthermore, by finding the individual marginal density functions, we obtain that the joint density is equal to the product of the individual marginal densities, which implies that $\|\mathbf{N}\|$ and $\boldsymbol{\theta}_{\mathbf{N}}$ are statistically independent. So are $\mathbf{U}$ and $\|\mathbf{N}\|$.

If $\Gamma(\gamma)$ and μ are used to denote the image of the path and the uniform probability measure on S^{p-1} respectively, define for $r \in [0, 1]$

$$\Gamma(\gamma)_{(r)} = \{u \in S^{p-1} : \sup(u^T v) > r \text{ for } v \in \Gamma(\gamma)\}. \qquad (5.5)$$

Recall (5.3). Since the supreme is no larger than one, we have $0 \le (\hat{\sigma}/\sigma)/\|\mathbf{N}\| \le 1/c$. Hence (5.3) is further equal to

$$1 - \int_0^{1/c} \mu\{[\Gamma(\gamma) \cup -\Gamma(\gamma)]_{(ct)}\} f_T(t)dt, \qquad (5.6)$$

where f_T denotes the density function of $T = (\hat{\sigma}/\sigma)/\|\mathbf{N}\|$ such that $pT^2 \sim F_{\nu,p}$, the F distribution with ν and p degrees of freedom.

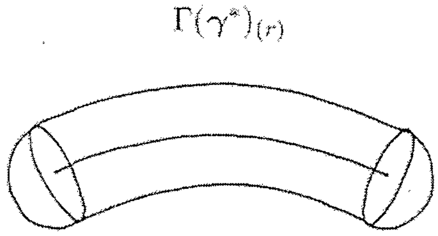

Figure 5.1: Tubular neighborhood of a path

The central part of Naiman (1986) is to find an upper bound of $\mu\{\Gamma(\gamma)_{(r)}\}$ so as to construct a conservative confidence band. Consider the case when $\Gamma(\gamma)$ can be piecewise approximated by great circular arcs using geometric inequalities. The great circular curve obtained after approximation can then be replaced by a curve of the same length on a single great circle by straightening out the curve at each point where the circular arcs are joined. Equivalently, if γ is replaced by γ^*, a path of equal length but whose image lies on a great circle, then the bound may be thought of as $\mu\{\Gamma(\gamma^*)_{(r)}\}$. Thus, a bound is obtained which depends only on the length of the path and consists of two terms. The first term is proportional to the length of the path corresponding to the points in the middle tubular part of $\Gamma(\gamma^*)_{(r)}$. The second term is the sum of the measures of two half spherical caps of angular radius $\cos^{-1} r$ corresponding to the points in the two half spherical ends of the tube $\Gamma(\gamma^*)_{(r)}$. Hence, the upper bound of $\mu\{\Gamma(\gamma)_{(r)}\}$ is given by

$$\begin{aligned}\mu\{\Gamma(\gamma)_{(r)}\} \leq \; & \min\{F_{p-2,2}[2(r^{-2}-1)/(p-2)] \times \Lambda(\gamma)/(2\pi) \\ & +F_{p-1,1}[(r^{-2}-1)/(p-1)]/2, 1\}. \qquad (5.7)\end{aligned}$$

The minimum used here is to avoid overlapping.

From (5.7), a lower bound for the coverage probability of the confidence band (5.2) is obtained as

$$\begin{aligned} 1 \; - \; & \int_0^{\frac{1}{c}} \min\{F_{p-2,2}[2((ct)^{-2}-1)/(p-2)] \times \Lambda(\gamma)/\pi \\ + \; & F_{p-1,1}[((ct)^{-2}-1)/(p-1)], 1\} f_T(t)dt, \qquad (5.8)\end{aligned}$$

where f_T is the density function of the random variable T, c is a critical value.

In a special case when $p = 2$ and $\mathrm{f}(x) = (1, x)^T$ where x belongs to a subset $\mathcal{X}$, the given model reduces to a usual simple linear regression model with a restricted predictor variable. Accordingly, the conservative confidence band becomes exact, because, for this special case, the path $\gamma(x)$ is already on the unit circle so that it is unnecessary to straighten it out. One may find $\mu\{\Gamma(\gamma)_{(r)}\}$ directly. Then the confidence level is

$$1 - \int_0^{\frac{1}{c}} \min\{\Lambda(\gamma)/\pi + F_{1,1}[(ct)^{-2} - 1], 1\} f_T(t)dt, \tag{5.9}$$

where $\Lambda(\gamma)$ is the length of the path.

It is of natural interest to show the equivalence between (5.9) and one of the computational formulae obtained in Chapter 3, which is used to calculate the critical values for the exact two-sided confidence bands for a simple linear regression.

Note that, in connection with (5.5), (5.6) can be written alternatively as

$$1 - \int_0^{\frac{1}{c}} \mathrm{P}\Big\{ \sup_{v \in \Gamma} |u^T v| > ct \Big\} f_T(t)dt. \tag{5.10}$$

By changing the variable of the integration, we have (5.10) further equal to

$$1 - \int_{\frac{c^2}{2}}^{\infty} \mathrm{P}\Big\{ \sup_{v \in \Gamma} |u^T v| > \frac{c}{\sqrt{2w}} \Big\} dF_{2,\nu}(w), \tag{5.11}$$

where $F_{2,\nu}$ stands for an F random variable with 2 and ν degrees of freedom.

Recall (3.34) and the definitions of u and v in (5.11), and we find that (5.11) is equivalent to formula (3.34). In connection with the equivalence of the formulae in Section 3.2, it can be concluded that the formula of Naiman's conservative method for the simple linear regression case is equivalent to that obtained by the exact method, e.g., formula (3.41).

5.2 Approximate confidence bands

Sun and Loader (1994) stated a method of constructing approximate $1-\alpha$ simultaneous confidence bands for a parametric or nonparametric function over a constrained predictor space. This method, which is in fact an approximation to the tube formula, can be applied to the multiple regression case, and is adaptable for a wide class of linear estimators. More details about the volume-of-tube formula, see, e.g., Loader (2004).

Consider the multiple regression model

$$y_i = f(\mathbf{x}_i) + \varepsilon, \tag{5.12}$$

where $\mathbf{x}_i, y_i, i = 1, \ldots, n$ are the observations, $\mathbf{x}_i \in \mathcal{R}^d$ is a vector of the predictor variables, $f(\cdot)$ is an unknown function which needs to be estimated based on the observations, and ε is the normally distributed random error with mean 0 and variance σ^2 which is assumed to be unknown. A linear estimator of the mean response $f(\mathbf{x})$ is given by

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{n} l_i(\mathbf{x}) y_i = \mathrm{l}(\mathbf{x})^T \mathbf{Y}, \tag{5.13}$$

where $\mathrm{l}(\mathbf{x}) = (l_1(\mathbf{x}), \ldots, l_n(\mathbf{x}))^T$ and $\mathbf{Y} = (y_1, \ldots, y_n)^T$.

A simultaneous confidence band for $f(\mathbf{x})$ over a subset $\mathcal{X}$ of the predictor space has the form given by

$$\{(\hat{f}(\mathbf{x}) - c\hat{\sigma}\|\mathrm{l}(\mathbf{x})\|, \hat{f}(\mathbf{x}) + c\hat{\sigma}\|\mathrm{l}(\mathbf{x})\|) : \mathbf{x} \in \mathcal{X}\}, \tag{5.14}$$

where c is a critical value and $\hat{\sigma}$ is the usual unbiased estimator of σ. If we assume the band (5.14) has $1-\alpha$ confidence level. Then, we have

$$\begin{aligned} 1-\alpha &= \inf_{f \in \mathcal{F}} \mathrm{P}_f\{\hat{f}(\mathbf{x}) - c\hat{\sigma}\|\mathrm{l}(\mathbf{x})\| \le f(\mathbf{x}) \\ &\qquad \le \hat{f}(\mathbf{x}) + c\hat{\sigma}\|\mathrm{l}(\mathbf{x})\|, \forall \mathbf{x} \in \mathcal{X}\}, \end{aligned} \tag{5.15}$$

where $\mathcal{F}$ is a wide suitable class of functions. Next, we evaluate the probability on the right-hand side of (5.15) in order to obtain a computational formula for calculating the critical value c.

A class $\mathcal{F}$, of natural interest, is a set of functions for which $\hat{f}(\mathbf{x})$ is an unbiased estimator, i.e.,

$$\mathcal{F} = \{f : f(\mathbf{x}) = \mathbf{l}(\mathbf{x})^T \boldsymbol{\mu}, \forall \mathbf{x}\},$$

where $\boldsymbol{\mu} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))^T$. In this case the probability that coverage fails for the band (5.14) has the following expressions

$$\begin{aligned}
\alpha &= \mathrm{P}\Big\{ \sup_{\mathbf{x}\in\mathcal{X}} \frac{|\hat{f}(\mathbf{x}) - f(\mathbf{x})|}{\|\mathbf{l}(\mathbf{x})\|} > c\hat{\sigma} \Big\} \\
&= \mathrm{P}\Big\{ \sup_{\mathbf{x}\in\mathcal{X}} \frac{|\mathbf{l}(\mathbf{x})^T \mathbf{Y} - \mathbf{l}(\mathbf{x})^T \boldsymbol{\mu}|}{\|\mathbf{l}(\mathbf{x})\|} > c\hat{\sigma} \Big\} \\
&= \mathrm{P}\Big\{ \sup_{\mathbf{x}\in\mathcal{X}} |\mathrm{T}(\mathbf{x})^T \boldsymbol{\varepsilon}| > c\hat{\sigma} \Big\} \\
&= \mathrm{P}\Big\{ \sup_{\mathbf{x}\in\mathcal{X}} \Big|\frac{\mathrm{T}(\mathbf{x})^T \boldsymbol{\varepsilon}}{\sigma}\Big| > \frac{c\hat{\sigma}}{\sigma} \Big\} \\
&= \mathrm{P}\Big\{ \sup_{\mathbf{x}\in\mathcal{X}} \Big|\frac{\mathrm{T}(\mathbf{x})^T \mathrm{N}}{\|\mathrm{N}\|}\Big| > c\frac{(\hat{\sigma}/\sigma)}{\|\mathrm{N}\|} \Big\} \\
&= \int_c^\infty \mathrm{P}\Big\{ \sup_{\mathbf{x}\in\mathcal{X}} |\mathrm{T}(\mathbf{x})^T \mathrm{U}| > \frac{c}{z} \Big\} g(z) dz, \qquad (5.16)
\end{aligned}$$

where $\mathrm{T}(\mathbf{x}) = \mathbf{l}(\mathbf{x})/\|\mathbf{l}(\mathbf{x})\|$, $\boldsymbol{\varepsilon} = \mathbf{Y} - \boldsymbol{\mu}$ is an n-dimensional vector of random errors, $g(z)$ is the density function of the random variable $Z = \|\mathrm{N}\|/(\hat{\sigma}/\sigma)$, $\mathrm{N} = \boldsymbol{\varepsilon}/\sigma$ has the $N_n(\mathbf{0}, \mathbf{l}(\mathbf{x})^{\mathrm{T}}\mathbf{l}(\mathbf{x}))$ distribution, and $\mathbf{U} = \mathrm{N}/\|\mathrm{N}\|$ is a unit vector on the surface of the unit sphere S^{n-1} and is independent of $\|\mathrm{N}\|$.

Letting $\mathcal{M} = \{\mathrm{T}(\mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$, the probability on the right-hand side of the last equality in (5.16) is simply the volume of a tubular neighborhood of $\mathcal{M} \cup -\mathcal{M}$ on the surface of S^{n-1}. Here, approximate formulae for one-dimensional and two-dimensional cases are given. Although this approximate method can be applied to high dimensional case by following the similar idea, lots of geometric constants are needed to be calculated. So we only consider the low dimensional cases, i.e., $d \leq 2$. The difficulty on the computation of the geometric constants is thought as the drawback of this method.

Assume the manifold $\mathcal{M}$ is the third order continuous with a positive critical radius. Suppose $\mathrm{T} : \mathcal{X} \longrightarrow \mathcal{M}$ is one-to-one, three times differentiable and there exists a vector $\boldsymbol{\lambda}$ such that $\boldsymbol{\lambda}^T \mathrm{T}(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{X}$ which ensures the tubes around $\mathcal{M}$ and $-\mathcal{M}$ do not intersect for sufficiently small radii.

Proposition 5.2.1 (One-dimensional) *Suppose $x \in [a, b]$. The length of $\mathcal{M}$ is $\kappa_0 = \int_a^b \|\mathrm{T}'(\mathrm{x})\|dx$, where $\mathrm{T}(\mathrm{x}) = \mathrm{l}(\mathrm{x})/\|\mathrm{l}(\mathrm{x})\|$ with $\mathrm{l}(\mathrm{x}) = X(X^T X)^{-1}\mathrm{x}$ for linear regression models, X is the design matrix. And $\nu\hat{\sigma}^2/\sigma^2 \sim \chi^2_\nu$. Then*

$$\alpha \approx \frac{\kappa_0}{\pi}(1 + \frac{c^2}{\nu})^{-\nu/2} + P\{|t_\nu| > c\}, \tag{5.17}$$

where the last term on the right-hand side of (5.17) is the probability of the absolute value of a t random variable with ν degrees of freedom larger than c.

Proposition 5.2.2 (Two-dimensional) *Suppose $\mathcal{X}$ is a rectangle in $\mathcal{R}^2$. Let κ_0 be the area of $\mathcal{M}$, ζ_0 be the length of the boundary of $\mathcal{M}$. Then*

$$\begin{aligned} \alpha \approx & \frac{\kappa_0}{\pi^{3/2}}\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\frac{c}{\sqrt{\nu}}(1 + \frac{c^2}{\nu})^{-(\nu+1)/2} \\ & + \frac{\zeta_0}{2\pi}(1 + \frac{c^2}{\nu})^{-\nu/2} + P\{|t_\nu| > c\}. \end{aligned} \tag{5.18}$$

For computing constants κ_0 and ζ_0, denote $\mathbf{T}_j(\mathrm{x}) = \partial \mathrm{T}(\mathrm{x})/\partial x_j$,j=1,2. Then

$$\kappa_0 = \int_{\mathcal{X}} det^{1/2}(A^T A)d\mathrm{x}, \tag{5.19}$$

$$\zeta_0 = \int_{\partial\mathcal{X}} det^{1/2}(\mathbf{A}_*{}^T \mathbf{A}_*), \tag{5.20}$$

where $A = (\mathbf{T}_1(\mathrm{x}), \mathbf{T}_2(\mathrm{x}))$, $\mathbf{A}_* = \mathbf{T}_1(\mathrm{x})$ or $\mathbf{T}_2(\mathrm{x})$.

5.3 Simulation-based confidence bands for a polynomial regression

Liu, Wynn and Hayter (2005) proposed the simulation-based method for constructing simultaneous confidence bands for an one-dimensional polynomial regression model with the only predictor variable restricted in an interval. Monte Carlo simulation is used to find an accurate approximation to the critical value of the confidence band when the number of simulations is set to be sufficiently large.

Consider the one-dimensional polynomial regression model

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5.21}$$

where $\mathbf{Y}_{n\times 1}$ is the vector of the observed responses, $X_{n\times p}$ is the full column-rank design matrix with the ith $(1 \le i \le n)$ row given by $(1, x_i, \ldots, x_i^{p-1})$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is the vector of unknown regression coefficients, and $\boldsymbol{\varepsilon}$ is the vector of independent and identically distributed normal random errors with mean 0 and variance σ^2, which is assumed to be unknown. Denote the maximum likelihood estimator of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, therefore, $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(X^TX)^{-1})$. Also, denote an unbiased estimator of σ^2 by $\hat{\sigma}^2$ so that $\nu\hat{\sigma}^2/\sigma^2 \sim \chi^2_\nu$. Moreover, $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent.

A hyperbolic-shape simultaneous confidence band for the mean response $\mathbf{x}^T\boldsymbol{\beta}$ over the restricted predictor space when $x \in (a, b)$ is given by

$$\mathbf{x}^T\boldsymbol{\beta} \in \mathbf{x}^T\hat{\boldsymbol{\beta}} \pm c\hat{\sigma}\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}}, \forall x \in (a, b), \tag{5.22}$$

where $\mathbf{x} = (1, x, \ldots, x^{p-1})$, c is a critical value such that the confidence band (5.22) has the confidence level equal to $1 - \alpha$. Alternatively, (5.22) can be arranged as

$$\sup_{a<x<b} \frac{|\mathbf{x}^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma|}{(\hat{\sigma}/\sigma)\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}}} \le c. \tag{5.23}$$

Define

$$T = \sup_{a<x<b} \frac{|\mathbf{x}^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma|}{(\hat{\sigma}/\sigma)\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}}}, \tag{5.24}$$

the confidence level of the band (5.22) is given by $\mathrm{P}\{T \le c\}$. The following procedure shows how to use Monte Carlo simulation method to approximate the critical value c.

Step 1 Generate $\mathrm{N} = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma \sim N_p(0, (X^{\mathbf{T}}X)^{-1})$.

Step 2 Generate $s = \hat{\sigma}/\sigma \sim \sqrt{\chi^2_\nu/\nu}$.

Step 3 Calculate T from (5.24). To find the supreme in (5.24), firstly find all the stationary points of

$$h(x) = \Big(\frac{|\mathbf{x}^T\mathbf{N}|}{\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}}}\Big)^2 = \frac{\mathbf{x}^T\mathbf{N}\mathbf{N}^T\mathbf{x}}{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}}$$

by solving $h'(x) = 0$, i.e.,

$$g(x) = \Big(\frac{d\mathbf{x}^T}{dx}\mathbf{N}\mathbf{N}^T\mathbf{x}\Big)\big[\mathbf{x}^T(X^TX)^{-1}\mathbf{x}\big] - \big(\mathbf{x}^T\mathbf{N}\mathbf{N}^T\mathbf{x}\big)\Big[\frac{d\mathbf{x}^T}{dx}(X^TX)^{-1}\mathbf{x}\Big] = 0.$$

Since $g(x)$ is a polynomial of order $4p-6$, it has at most $4p-6$ zero points. If they are denoted by $x_1, \ldots, x_q$, from (5.24) we have

$$T = \max\{\sqrt{h(a)}, \sqrt{h(b)}, \max_{1\le i\le q; x_i\in(a,b)} \sqrt{h(x_i)}\}/s.$$

Step 4 Simulate R independent replicates of T, say, $T_1, \ldots, T_R$, and use the $[(1-\alpha)R]$th largest T_i as an approximation of c, denoted by $\hat{c}$.

The base of this approach is that the sample $100(1-\alpha)$ percentile $\hat{c}$ converges almost surely to the population $100(1-\alpha)$ percentile c when the number of simulations R goes to infinity. Furthermore, to gauge the accuracy of $\hat{c}$, it is useful to estimate its standard error. It is known that, under certain regularity conditions, $\hat{c}$ is asymptotically normal with mean c and standard error

$$s.e. = \sqrt{\frac{\alpha(1-\alpha)}{RG^2(c)}}, \tag{5.25}$$

where $G(c)$ is the density function of T evaluated at c (see, e.g., Serfling, 1980). And $G(c)$ may be approximated by the kernel density estimator

$$G(c) = \frac{1}{Rd\sqrt{2\pi}}\sum_{i=1}^{R} e^{-[(\hat{c}-T_i)/d]^2/2},$$

where T_i is the ith simulated value and d is the smoothing parameter. We usually set $d = 0.1, 0.01, 0.001$.

5.4 Simulation-based confidence bands for a multiple linear regression

Liu, Jamshidian, Zhang and Donnelly (2005) presented a method of constructing simultaneous confidence bands for a normal-error multiple linear regression model based on Monte Carlo simulation procedure. The confidence bands constructed via this method have hyperbolic shape and can be applied to a model with any number of predictor variables.

Consider the multiple linear regression model

$$\mathbf{Y} = X\boldsymbol{\beta} + \varepsilon,$$

where $X_{n\times p}$ is the design matrix with the first column given by $(1,\ldots,1)^T$ and the jth $(2 \le j \le p)$ column given by $(x_{1,j-1},\ldots,x_{n,j-1})^T$. Inferences on estimators of unknown parameters $\boldsymbol{\beta}$ and σ^2 can be obtained as usual.

It is of interest to construct a simultaneous confidence band on the most popular rectangular region $\mathcal{X}$ of the predictor space, which is of the form

$$\mathcal{X} = \{(x_1,\ldots,x_{p-1})^T : a_i \le x_i \le b_i, i = 1,\ldots,p-1\}, \tag{5.26}$$

where $-\infty \le a_i < b_i \le \infty, i = 1,\ldots,p-1$ are given constants. The central task is to find an appropriate critical value c such that the confidence band has the confidence level equal to a preassigned $1-\alpha$.

Note that the confidence level of the band is given by $\mathrm{P}\{T < c\}$, where

$$T = \sup_{x_i\in[a_i,b_i],i=1,\ldots,p-1} \frac{|\mathbf{x}^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})/\sigma|}{(\hat{\sigma}/\sigma)\sqrt{\mathbf{x}^T(X^TX)^{-1}\mathbf{x}}}. \tag{5.27}$$

The distribution of T depends on the design matrix and the intervals (a_i, b_i) in a complicated manner. This makes a challenge to derive the distribution function of T directly. In such a case, it is motivated to introduce a simulation-based method to find an approximation to the critical value c, say $\hat{c}$, which can be as accurate as one wants by simulating a sufficiently large number of T's.

It is clear from (5.27) that the calculation of T is in fact an optimization problem. Consequently, our analysis focuses on the optimization algorithm. Let P be a $p \times p$ non-singular matrix such that $(X^T X)^{-1} = P^T P$. Then generate one $\mathbf{N} = (P^T)^{-1}(\hat{\beta} - \beta)/\sigma \sim N_p(0, I)$ and one independent $\hat{\sigma}/\sigma \sim \sqrt{\chi^2_\nu/\nu}$. Therefore, T becomes

$$T = Q\frac{||\mathbf{N}||}{(\hat{\sigma}/\sigma)}, \tag{5.28}$$

where

$$Q = \sup_{x_i \in [a_i, b_i], i=1,\ldots,p-1} \frac{|(P\mathbf{x})^T \mathbf{N}|}{||P\mathbf{x}||\,||\mathbf{N}||}.$$

Accordingly, the optimization problem of T transforms to the optimization of Q which involves the maximization of a p-variate function over the given rectangular region of the predictor space. Two methods were included in Liu, Jamshidian, Zhang and Donnelly (2005) to solve such a maximization problem. They are the branching method and the active set method respectively. T can be obtained after Q is ready.

As stated in Section 5.3, we simulate R replicates of the random variable T, and set the $[(1-\alpha)R]$th largest simulated value $\hat{c}$ as an approximate of the critical value c. Also, one may estimate the standard error of $\hat{c}$ using (5.25) to gauge its accuracy.

5.5 Comparisons

For the methods of constructing confidence bands introduced in this chapter, we are interested in comparing them in terms of the critical value to have a general view on the goodness of each. All the critical values in this section are calculated using MATLAB programmes.

5.5.1 For simple linear regression

We start with the comparison for a simple linear regression model. As already pointed out, for simple linear regression case, Naiman's method turns to be exact. So the methods in our first comparison include: the exact method, the approximate method of Sun and Loader (1994), and the simulation-based method of Liu, Wynn and Hayter (2005).

Note that, all the methods depend on the design matrix, the restricted interval for the only predictor variable and the confidence level. However, one may go further to consider the nature of the methods.

For the exact method, we have the computational formula for the critical value given by

$$1-\alpha=\frac{2\theta^*}{\pi}F_{2,\nu}(\frac{c^2}{2})+\frac{2}{\pi}\int_0^{\frac{\pi}{2}-\theta^*}F_{2,\nu}(\frac{c^2}{2\cos^2\theta})\mathrm{d}\theta, \tag{5.29}$$

where $F_{2,\nu}$ stands for the F cumulative distribution function with 2 and $\nu=n-2$ degrees of freedom, and θ^* can be found in the following way: define $\mathbf{a}=(1,a)^T$, $\mathbf{b}=(1,b)^T$ where a,b are the lower and upper bounds of the restricted interval, then we have

$$\theta^*=\frac{1}{2}\arccos\frac{\mathbf{a}^T(X^TX)^{-1}\mathbf{b}}{(\mathbf{a}^T(X^TX)^{-1}\mathbf{a}\cdot\mathbf{b}^T(X^TX)^{-1}\mathbf{b})^{1/2}}. \tag{5.30}$$

It is clear that the critical value depends on the angle θ^*, the degree of freedom ν and the given confidence level $1-\alpha$, where θ^* is half the angle between $P\mathbf{a}$ and $P\mathbf{b}$ with P consistently defined as before.

Similar argument can be applied to the approximate method, the key of which is to compute the length of the path on the surface of the unit sphere in $\mathcal{R}^n$. So we are interested in finding the relationship between the length of the path and the angle θ^*.

For linear regression models, vector $\mathrm{l}(\mathrm{x})$ in the approximate method has the explicit form given by

$$\mathrm{l}(\mathrm{x})=X(X^TX)^{-1}\mathrm{x}=XP^TP\mathrm{x},$$

where X is the design matrix, x is the vector of the covariates, and

$$\|\mathrm{l}(\mathrm{x})\|=\Big[\mathrm{l}(\mathrm{x})^T\mathrm{l}(\mathrm{x})\Big]^{\frac{1}{2}}=\Big[\mathrm{x}^T(X^TX)^{-1}\mathrm{x}\Big]^{\frac{1}{2}}=\|P\mathrm{x}\|.$$

Then we have

$$\mathbf{T}(\mathrm{x}) = \frac{\mathrm{l}(\mathrm{x})}{\|\mathrm{l}(\mathrm{x})\|} = \frac{XP^T \cdot P\mathrm{x}}{\|P\mathrm{x}\|} = XP^T \cdot \gamma(\mathrm{x})$$

by studying (5.4) for Naiman's method. Furthermore, we have

$$\begin{aligned}
\|\mathrm{T}'(\mathrm{x})\|^2 &= \Big[\mathrm{T}'(\mathrm{x})\Big]^T \Big[\mathrm{T}'(\mathrm{x})\Big] \\
&= \Big[\gamma'(\mathrm{x})\Big]^T PX^T \cdot XP^T \Big[\gamma'(\mathrm{x})\Big] \\
&= \Big[\gamma'(\mathrm{x})\Big]^T \Big[P(P^T P)^{-1} P^T\Big] \Big[\gamma'(\mathrm{x})\Big] \\
&= \Big[\gamma'(\mathrm{x})\Big]^T I_p \Big[\gamma'(\mathrm{x})\Big] = \|\gamma'(\mathrm{x})\|^2,
\end{aligned}$$

which implies that

$$\|\mathrm{T}'(\mathrm{x})\| = \|\gamma'(\mathrm{x})\|. \tag{5.31}$$

Thus, by assuming the only predictor variable $x \in [a, b]$, since the length of the path

$$\Lambda(\gamma) = \int_a^b \|\gamma'(\mathrm{x})\| dx = \int_a^b \|\mathrm{T}'(\mathrm{x})\| dx = \kappa_0, \tag{5.32}$$

then we obtain the equivalence between the length of the path in Naiman's method and that in the approximate method of Sun and Loader (1994).

In particular for simple linear regression case, the path in Naiman's method is on the unit circle, which, in connection with the fact that $2\theta^*$ is equal to the angle between the two unit vectors starting from the origin and pointing to the two ends of the path, implies that $2\theta^* = \Lambda(\gamma) = \kappa_0$.

Therefore, formula (5.17) becomes

$$\alpha = \frac{2\theta^*}{\pi}(1 + \frac{c^2}{\nu})^{-\nu/2} + \mathrm{P}\{|t_\nu| > c\}, \tag{5.33}$$

where t_ν is a t random variable with ν degrees of freedom. Clearly, the critical value depends on θ^*, ν, $1 - \alpha$ as well.

For the simulation-based method, a suitable manipulation simplifies the computation of T defined in (5.24). Define $\mathbf{U} = (U_1, U_2)^T = \mathbf{N}/\|\mathbf{N}\|$, where $\mathbf{N} = (P^T)^{-1}(\hat{\beta} - \beta)/\sigma$. Let u_1 and u_2 be the generated values of random variables U_1 and U_2. Then, under the definition of $\phi^* = 2\theta^*$, T can be calculated in the following way:

$$T = \begin{cases} \frac{\|\mathbf{N}\|}{(\hat{\sigma}/\sigma)} & \text{if } |u_1| \geq \cos\frac{\phi^*}{2}, \\ \frac{\|\mathbf{N}\|}{(\hat{\sigma}/\sigma)}|u_1 \cos\frac{\phi^*}{2} + u_2 \sin\frac{\phi^*}{2}| & \text{if } 0 \leq u_1 < \cos\frac{\phi^*}{2}\ ,\ u_2 > \sin\frac{\phi^*}{2} \\ & \text{or } -\cos\frac{\phi^*}{2} < u_1 < 0\ ,\ u_2 < -\sin\frac{\phi^*}{2}, \\ \frac{\|\mathbf{N}\|}{(\hat{\sigma}/\sigma)}|u_1 \cos\frac{\phi^*}{2} - u_2 \sin\frac{\phi^*}{2}| & \text{otherwise.} \end{cases}$$

By following the procedure stated in Section 5.3, an approximation of the critical value can be found. Also, one may calculate the standard error accordingly. Overall, all these methods depend on θ^*, ν and $1-\alpha$.

Now, we design the levels for these three factors. Since $\theta^* \in [0, \pi/2]$, set $\theta^* = 0.0, 0.1, 0.2, \cdots, 1.3, 1.4, 1.5, 1.57$. Set the degree of freedom $\nu = 2, 4, 6, 8, 10, 15, 20, 30, 40, 60$, from small to large, to see how this factor affects the critical value. In addition, the three most popular confidence levels $90\%, 95\%, 99\%$ are used. Tables 5.1-5.10 contain the critical values computed based on these methods. The simulation results are based on 100,000 simulations. Results with a star behind in the tables are based on 200,000 simulations in order to make the distinction more clear.

From the results, we can draw some conclusions. For each method, the critical value increases with the angle θ^* and the confidence level, and decreases with the degree of freedom. The critical values based on the approximate method are generally larger than those of the exact method, good enough when θ^* takes small values, but being worse and worse as θ^* goes large. This trend becomes less and less evident as ν goes large, but the gap still exists clearly. The simulation-based method computes as good critical values as the exact method. The difference between the critical values of the simulation-based method and those of the approximate method is basically at the second or third decimal place, increasing with the confidence level and decreasing with the degree of freedom.

d.f. = 2

cl	θ*	cv_ext	cv_nai	cv_app	cv_simu	s.e.	
	0.00	2.9200	2.9200	2.9200	2.9103	0.0153	
	0.10	3.1200	3.1200	3.1208	3.1125	0.0161	
	0.20	3.3055	3.3055	3.3114	3.2976	0.0210	
	0.30	3.4746	3.4746	3.4927	3.4589	0.0128	*
	0.40	3.6260	3.6260	3.6660	3.6050	0.0132	*
	0.50	3.7592	3.7592	3.8321	3.7489	0.0137	*
	0.60	3.8742	3.8742	3.9918	3.8454	0.0228	
	0.70	3.9712	3.9712	4.1458	3.9522	0.0190	
90%	0.80	4.0509	4.0509	4.2945	4.0505	0.0154	*
	0.90	4.1142	4.1140	4.4385	4.1056	0.0211	
	1.00	4.1622	4.1622	4.5782	4.1489	0.0215	
	1.10	4.1969	4.1969	4.7140	4.1971	0.0147	*
	1.20	4.2200	4.2200	4.8459	4.2068	0.0190	
	1.30	4.2337	4.2337	4.9746	4.1875	0.0189	
	1.40	4.2404	4.2404	5.1000	4.2238	0.0223	
	1.50	4.2425	4.2425	5.2226	4.2422	0.0202	
	1.57	4.2427	4.2427	5.3067	4.2432	0.0159	*
	0.00	4.3027	4.3026	4.3027	4.2814	0.0302	
	0.10	4.5816	4.5816	4.5828	4.5012	0.0299	
	0.20	4.8403	4.8403	4.8478	4.8374	0.0344	
	0.30	5.0765	5.0765	5.1000	5.0591	0.0306	*
	0.40	5.2887	5.2887	5.3408	5.2892	0.0453	
	0.50	5.4761	5.4761	5.5718	5.4712	0.0331	
	0.60	5.6386	5.6386	5.7937	5.6107	0.0281	*
	0.70	5.7760	5.7758	6.0078	5.7418	0.0393	
95%	0.80	5.8892	5.8892	6.2149	5.7967	0.0402	
	0.90	5.9794	5.9794	6.4152	5.9549	0.0455	
	1.00	6.0486	6.0484	6.6097	6.0284	0.0317	*
	1.10	6.0984	6.0984	6.7988	6.0639	0.0453	
	1.20	6.1317	6.1317	6.9828	6.1348	0.0553	
	1.30	6.1516	6.1516	7.1621	6.1127	0.0521	
	1.40	6.1611	6.1611	7.3372	6.1565	0.0310	*
	1.50	6.1641	6.1641	7.5082	6.1580	0.0297	*
	1.57	6.1643	6.1643	7.6256	6.1350	0.0479	
	0.00	9.9249	9.9247	9.9249	9.8140	0.0662	*
	0.10	10.5420	10.5420	10.5441	10.6691	0.1004	*
	0.20	11.1141	11.1141	11.1295	11.1316	0.1312	
	0.30	11.6370	11.6370	11.6858	11.5289	0.0833	
	0.40	12.1081	12.1079	12.2173	12.0786	0.1430	*
	0.50	12.5250	12.5250	12.7267	12.5740	0.4441	
	0.60	12.8872	12.8872	13.2167	12.7368	0.2089	
	0.70	13.1949	13.1947	13.6891	13.2957	0.1347	
99%	0.80	13.4491	13.4491	14.1460	13.0626	0.1170	
	0.90	13.6525	13.6523	14.5865	13.1477	0.2324	
	1.00	13.8082	13.8082	15.0182	13.7085	0.1167	*
	1.10	13.9213	13.9211	15.4359	14.2853	0.3344	
	1.20	13.9920	13.9920	15.8425	14.0068	0.2436	
	1.30	14.0419	14.0419	16.2391	14.2307	0.2019	
	1.40	14.0639	14.0637	16.6263	13.8603	0.1011	*
	1.50	14.0707	14.0705	17.0047	14.0654	0.2143	
	1.57	14.0713	14.0713	17.2645	14.0603	0.1750	

Table 5.1: Critical values for simple linear regression

d.f. = 4

cl	θ^*	cv_ext	cv_nai	cv_app	cv_simu	s.e.	
90%	0.00	2.1319	2.1317	2.1319	2.1250	0.0057	*
	0.10	2.2505	2.2503	2.2505	2.2515	0.0091	
	0.20	2.3590	2.3590	2.3592	2.3604	0.0092	
	0.30	2.4583	2.4583	2.4594	2.4575	0.0090	
	0.40	2.5485	2.5485	2.5523	2.5432	0.0065	*
	0.50	2.6293	2.6293	2.6290	2.6265	0.0066	*
	0.60	2.7006	2.7006	2.7202	2.6980	0.0067	*
	0.70	2.7618	2.7618	2.7967	2.7504	0.0063	*
	0.80	2.8131	2.8131	2.8688	2.8160	0.0077	*
	0.90	2.8545	2.8545	2.9372	2.8503	0.0100	
	1.00	2.8865	2.8865	3.0022	2.8654	0.0101	
	1.10	2.9097	2.9097	3.0642	2.9063	0.0097	
	1.20	2.9254	2.9254	3.1234	2.9224	0.0108	
	1.30	2.9349	2.9349	3.1803	2.9360	0.0074	*
	1.40	2.9395	2.9393	3.2348	2.9435	0.0072	*
	1.50	2.9408	2.9408	3.2872	2.9364	0.0111	
	1.57	2.9410	2.9410	3.3227	2.9324	0.0074	*
95%	0.00	2.7765	2.7765	2.7765	2.7776	0.0136	
	0.10	2.9147	2.9145	2.9147	2.9079	0.0132	
	0.20	3.0403	3.0403	3.0405	3.0137	0.0148	
	0.30	3.1549	3.1549	3.1559	3.1362	0.0151	
	0.40	3.2592	3.2592	3.2626	3.2580	0.0134	
	0.50	3.3532	3.3532	3.3619	3.3510	0.0169	
	0.60	3.4367	3.4367	3.4550	3.4397	0.0146	
	0.70	3.5091	3.5091	3.5423	3.5132	0.0100	*
	0.80	3.5703	3.5701	3.6248	3.5739	0.0160	
	0.90	3.6201	3.6201	3.7030	3.6169	0.0108	*
	1.00	3.6591	3.6591	3.7773	3.6472	0.0172	
	1.10	3.6877	3.6877	3.8463	3.6765	0.0148	
	1.20	3.7074	3.7074	3.9161	3.7095	0.0110	*
	1.30	3.7190	3.7190	3.9809	3.7099	0.0168	
	1.40	3.7247	3.7247	4.0433	3.7119	0.0193	
	1.50	3.7266	3.7266	4.1033	3.7181	0.0148	
	1.57	3.7268	3.7268	4.1441	3.7210	0.0112	*
99%	0.00	4.6041	4.6040	4.6041	4.5947	0.0305	*
	0.10	4.8056	4.8056	4.8056	4.8122	0.0425	
	0.20	4.9873	4.9873	4.9875	4.9801	0.0385	
	0.30	5.1526	5.1526	5.1536	5.1292	0.0316	
	0.40	5.3030	5.3030	5.3065	5.2951	0.0321	*
	0.50	5.4391	5.4391	5.4485	5.4124	0.0357	*
	0.60	5.5608	5.5608	5.5812	5.5327	0.0420	
	0.70	5.6677	5.6675	5.7058	5.6595	0.0279	*
	0.80	5.7588	5.7588	5.8235	5.7466	0.0320	*
	0.90	5.8343	5.8343	5.9348	5.8554	0.0351	*
	1.00	5.8940	5.8940	6.0408	5.8222	0.0315	*
	1.10	5.9386	5.9384	6.1418	5.8970	0.0347	*
	1.20	5.9691	5.9691	6.2385	5.9874	0.0324	*
	1.30	5.9876	5.9876	6.3311	5.9580	0.0445	
	1.40	5.9970	5.9968	6.4202	5.8860	0.0428	
	1.50	5.9998	5.9998	6.5058	5.9738	0.0499	
	1.57	6.0000	6.0000	6.5639	6.0047	0.0339	*

Table 5.2: Critical values for simple linear regression

d.f. = 6

cl	θ*	cv_ext	cv_nai	cv_app	cv_simu	s.e.	
	0.00	1.9432	1.9432	1.9432	1.9493	0.0050	*
	0.10	2.0440	2.0440	2.0440	2.0332	0.0067	
	0.20	2.1354	2.1354	2.1354	2.1297	0.0068	
	0.30	2.2183	2.2183	2.2185	2.2007	0.0071	
	0.40	2.2938	2.2938	2.2947	2.2944	0.0048	*
	0.50	2.3618	2.3618	2.3653	2.3559	0.0072	
	0.60	2.4225	2.4225	2.4305	2.4215	0.0076	
	0.70	2.4749	2.4749	2.4913	2.4669	0.0081	
90%	0.80	2.5193	2.5193	2.5483	2.5155	0.0074	
	0.90	2.5553	2.5553	2.6019	2.5490	0.0077	
	1.00	2.5836	2.5836	2.6524	2.5851	0.0072	
	1.10	2.6041	2.6041	2.7000	2.6018	0.0077	
	1.20	2.6181	2.6181	2.7454	2.6232	0.0056	*
	1.30	2.6264	2.6264	2.7887	2.6231	0.0054	*
	1.40	2.6305	2.6305	2.8299	2.6348	0.0079	
	1.50	2.6318	2.6318	2.8693	2.6318	0.0076	
	1.57	2.6318	2.6318	2.8958	2.6295	0.0075	
	0.00	2.4469	2.4469	2.4469	2.4509	0.0071	*
	0.10	2.5578	2.5578	2.5578	2.5564	0.0105	
	0.20	2.6573	2.6573	2.6573	2.6526	0.0116	
	0.30	2.7471	2.7471	2.7473	2.7410	0.0079	*
	0.40	2.8287	2.8287	2.8295	2.8263	0.0074	*
	0.50	2.9027	2.9025	2.9050	2.8947	0.0098	
	0.60	2.9686	2.9686	2.9747	2.9650	0.0111	
	0.70	3.0268	3.0268	3.0398	3.0231	0.0075	*
95%	0.80	3.0765	3.0765	3.1006	3.0686	0.0078	*
	0.90	3.1175	3.1175	3.1578	3.1098	0.0114	
	1.00	3.1499	3.1499	3.2115	3.1458	0.0120	
	1.10	3.1742	3.1742	3.2626	3.1675	0.0113	
	1.20	3.1907	3.1906	3.3108	3.1882	0.0130	
	1.30	3.2007	3.2007	3.3570	3.1998	0.0088	*
	1.40	3.2056	3.2056	3.4008	3.1871	0.0107	
	1.50	3.2071	3.2071	3.4428	3.2046	0.0103	
	1.57	3.2073	3.2073	3.4712	3.2002	0.0113	
	0.00	3.7074	3.7074	3.7074	3.6413	0.0252	
	0.10	3.8484	3.8484	3.8484	3.8314	0.0267	
	0.20	3.9731	3.9731	3.9731	3.9349	0.0234	
	0.30	4.0848	4.0847	4.0848	4.1056	0.0340	
	0.40	4.1857	4.1857	4.1861	4.1426	0.0319	
	0.50	4.2772	4.2770	4.2787	4.2697	0.0387	
	0.60	4.3597	4.3596	4.3641	4.3371	0.0239	
	0.70	4.4331	4.4331	4.4436	4.4596	0.0198	
99%	0.80	4.4972	4.4972	4.5176	4.5096	0.0200	*
	0.90	4.5511	4.5511	4.5872	4.5135	0.0295	
	1.00	4.5946	4.5946	4.6528	4.5632	0.0298	
	1.10	4.6276	4.6276	4.7147	4.6103	0.0213	*
	1.20	4.6507	4.6507	4.7736	4.6337	0.0260	
	1.30	4.6648	4.6648	4.8297	4.6455	0.0177	*
	1.40	4.6718	4.6718	4.8830	4.6696	0.0302	
	1.50	4.6741	4.6741	4.9341	4.6181	0.0279	
	1.57	4.6743	4.6743	4.9686	4.6226	0.0380	

Table 5.3: Critical values for simple linear regression

d.f. = 8

cl	θ*	cv_ext	cv_nai	cv_app	cv_simu	s.e.	
	0.00	1.8595	1.8595	1.8595	1.8595	0.0042	*
	0.10	1.9529	1.9527	1.9529	1.9519	0.0042	*
	0.20	2.0368	2.0368	2.0368	2.0312	0.0067	
	0.30	2.1127	2.1127	2.1127	2.1130	0.0042	*
	0.40	2.1817	2.1817	2.1821	2.1706	0.0066	
	0.50	2.2440	2.2440	2.2456	2.2442	0.0066	
	0.60	2.2997	2.2995	2.3043	2.2969	0.0066	
	0.70	2.3483	2.3481	2.3586	2.3432	0.0067	
90%	0.80	2.3895	2.3895	2.4093	2.3910	0.0066	
	0.90	2.4232	2.4232	2.4568	2.4299	0.0065	
	1.00	2.4495	2.4495	2.5012	2.4451	0.0046	*
	1.10	2.4690	2.4690	2.5431	2.4628	0.0066	
	1.20	2.4821	2.4821	2.5830	2.4847	0.0065	
	1.30	2.4901	2.4901	2.6205	2.4850	0.0069	
	1.40	2.4940	2.4940	2.6564	2.4916	0.0067	
	1.50	2.4951	2.4951	2.6905	2.4890	0.0048	*
	1.57	2.4953	2.4953	2.7136	2.4859	0.0066	
	0.00	2.3060	2.3060	2.3060	2.3095	0.0062	*
	0.10	2.4059	2.4059	2.4059	2.4108	0.0062	*
	0.20	2.4945	2.4945	2.4945	2.4903	0.0087	
	0.30	2.5742	2.5742	2.5742	2.5867	0.0095	
	0.40	2.6463	2.6463	2.6465	2.6494	0.0066	*
	0.50	2.7115	2.7115	2.7124	2.7051	0.0095	
	0.60	2.7700	2.7700	2.7731	2.7703	0.0091	
	0.70	2.8220	2.8219	2.8291	2.8165	0.0062	*
95%	0.80	2.8667	2.8667	2.8813	2.8604	0.0073	*
	0.90	2.9040	2.9038	2.9303	2.9002	0.0070	*
	1.00	2.9336	2.9336	2.9761	2.9251	0.0090	
	1.10	2.9557	2.9557	3.0192	2.9374	0.0085	
	1.20	2.9709	2.9709	3.0602	2.9540	0.0097	
	1.30	2.9801	2.9801	3.0989	2.9768	0.0083	
	1.40	2.9847	2.9847	3.1356	2.9777	0.0086	
	1.50	2.9862	2.9862	3.1707	2.9708	0.0068	*
	1.57	2.9864	2.9864	3.1944	2.9846	0.0073	*
	0.00	3.3555	3.3553	3.3555	3.3605	0.0144	*
	0.10	3.4740	3.4740	3.4740	3.4398	0.0213	
	0.20	3.5776	3.5776	3.5776	3.5720	0.0157	*
	0.30	3.6694	3.6694	3.6694	3.6514	0.0144	*
	0.40	3.7518	3.7518	3.7520	3.7517	0.0221	
	0.50	3.8263	3.8263	3.8267	3.7964	0.0205	
	0.60	3.8936	3.8936	3.8953	3.8705	0.0143	*
	0.70	3.9543	3.9541	3.9586	3.9431	0.0203	
99%	0.80	4.0074	4.0073	4.0174	4.0125	0.0167	*
	0.90	4.0528	4.0526	4.0721	4.0323	0.0169	*
	1.00	4.0898	4.0898	4.1235	4.0885	0.0257	
	1.10	4.1182	4.1182	4.1720	4.1171	0.0233	
	1.20	4.1382	4.1382	4.2177	4.1273	0.0180	*
	1.30	4.1506	4.1506	4.2610	4.1396	0.0213	
	1.40	4.1569	4.1569	4.3022	4.1405	0.0161	*
	1.50	4.1590	4.1590	4.3416	4.1434	0.0141	*
	1.57	4.1592	4.1592	4.3679	4.1477	0.0178	

Table 5.4: Critical values for simple linear regression

d.f. = 10

cl	θ^*	cv_ext	cv_nai	cv_app	cv_simu	s.e.	
	0.00	1.8124	1.8124	1.8124	1.8109	0.0041	*
	0.10	1.9016	1.9016	1.9016	1.8993	0.0044	*
	0.20	1.9815	1.9813	1.9815	1.9797	0.0060	
	0.30	2.0534	2.0534	2.0536	2.0490	0.0060	
	0.40	2.1188	2.1188	2.1190	2.1180	0.0058	
	0.50	2.1779	2.1777	2.1788	2.1791	0.0041	*
	0.60	2.2307	2.2307	2.2337	2.2320	0.0041	*
	0.70	2.2770	2.2770	2.2846	2.2773	0.0060	
90%	0.80	2.3165	2.3165	2.3319	2.3120	0.0056	
	0.90	2.3489	2.3489	2.3759	2.3485	0.0064	
	1.00	2.3744	2.3744	2.4173	2.3750	0.0064	
	1.10	2.3931	2.3931	2.4562	2.3852	0.0064	
	1.20	2.4059	2.4059	2.4928	2.3958	0.0059	
	1.30	2.4135	2.4135	2.5275	2.4134	0.0062	
	1.40	2.4171	2.4171	2.5605	2.4140	0.0058	
	1.50	2.4183	2.4183	2.5918	2.4135	0.0067	
	1.57	2.4185	2.4185	2.6129	2.4177	0.0063	
	0.00	2.2282	2.2280	2.2282	2.2182	0.0077	
	0.10	2.3220	2.3220	2.3220	2.3101	0.0086	
	0.20	2.4049	2.4049	2.4049	2.4042	0.0083	
	0.30	2.4791	2.4791	2.4791	2.4790	0.0084	
	0.40	2.5458	2.5458	2.5460	2.5236	0.0085	
	0.50	2.6064	2.6062	2.6068	2.5976	0.0077	
	0.60	2.6608	2.6608	2.6627	2.6513	0.0087	
	0.70	2.7092	2.7092	2.7140	2.7092	0.0093	
95%	0.80	2.7511	2.7511	2.7616	2.7382	0.0084	
	0.90	2.7862	2.7862	2.8062	2.7853	0.0092	
	1.00	2.8142	2.8142	2.8478	2.8124	0.0090	
	1.10	2.8354	2.8352	2.8869	2.8332	0.0087	
	1.20	2.8499	2.8499	2.9237	2.8396	0.0087	
	1.30	2.8586	2.8586	2.9585	2.8557	0.0086	
	1.40	2.8630	2.8630	2.9917	2.8585	0.0064	*
	1.50	2.8644	2.8644	3.0232	2.8607	0.0060	*
	1.57	2.8646	2.8646	3.0443	2.8595	0.0088	
	0.00	3.1692	3.1692	3.1692	3.1669	0.0194	
	0.10	3.2765	3.2763	3.2765	3.2794	0.0123	*
	0.20	3.3692	3.3692	3.3692	3.3656	0.0190	
	0.30	3.4510	3.4510	3.4510	3.4360	0.0146	
	0.40	3.5238	3.5238	3.5240	3.5137	0.0165	
	0.50	3.5896	3.5896	3.5898	3.5688	0.0218	
	0.60	3.6490	3.6490	3.6498	3.6539	0.0154	*
	0.70	3.7026	3.7026	3.7051	3.7003	0.0196	
99%	0.80	3.7501	3.7501	3.7562	3.7300	0.0132	*
	0.90	3.7909	3.7909	3.8036	3.7711	0.0214	
	1.00	3.8244	3.8244	3.8479	3.8110	0.0190	
	1.10	3.8505	3.8505	3.8896	3.8230	0.0158	*
	1.20	3.8690	3.8688	3.9289	3.8625	0.0136	*
	1.30	3.8805	3.8805	3.9661	3.8943	0.0196	
	1.40	3.8862	3.8862	4.0013	3.8513	0.0140	*
	1.50	3.8881	3.8881	4.0347	3.8726	0.0168	
	1.57	3.8883	3.8883	4.0574	3.8596	0.0141	*

Table 5.5: Critical values for simple linear regression

d.f. = 15

cl	θ^*	cv_ext	cv_nai	cv_app	cv_simu	s.e.	
90%	0.00	1.7531	1.7529	1.7531	1.7534	0.0052	
	0.10	1.8370	1.8370	1.8370	1.8343	0.0057	
	0.20	1.9119	1.9117	1.9119	1.9117	0.0038	*
	0.30	1.9790	1.9790	1.9790	1.9792	0.0056	
	0.40	2.0397	2.0397	2.0397	2.0383	0.0052	
	0.50	2.0946	2.0946	2.0949	2.0905	0.0054	
	0.60	2.1439	2.1439	2.1455	2.1505	0.0053	
	0.70	2.1874	2.1874	2.1922	2.1846	0.0040	*
	0.80	2.2246	2.2246	2.2353	2.2218	0.0040	*
	0.90	2.2553	2.2553	2.2753	2.2532	0.0036	*
	1.00	2.2795	2.2795	2.3127	2.2612	0.0057	
	1.10	2.2974	2.2974	2.3477	2.2955	0.0041	*
	1.20	2.3096	2.3096	2.3807	2.3078	0.0051	
	1.30	2.3169	2.3169	2.4118	2.3174	0.0039	*
	1.40	2.3205	2.3205	2.4413	2.3193	0.0057	
	1.50	2.3216	2.3216	2.4692	2.3129	0.0055	
	1.57	2.3216	2.3216	2.4879	2.3157	0.0058	
95%	0.00	2.1314	2.1314	2.1314	2.1358	0.0056	*
	0.10	2.2179	2.2179	2.2179	2.2154	0.0075	
	0.20	2.2940	2.2940	2.2940	2.2843	0.0075	
	0.30	2.3615	2.3615	2.3615	2.3491	0.0063	
	0.40	2.4219	2.4219	2.4219	2.4201	0.0069	
	0.50	2.4764	2.4764	2.4766	2.4662	0.0049	*
	0.60	2.5256	2.5256	2.5264	2.5222	0.0056	*
	0.70	2.5696	2.5696	2.5723	2.5621	0.0073	
	0.80	2.6081	2.6081	2.6144	2.6142	0.0049	*
	0.90	2.6406	2.6406	2.6537	2.6263	0.0078	
	1.00	2.6665	2.6665	2.6903	2.6668	0.0071	
	1.10	2.6863	2.6863	2.7246	2.6893	0.0051	*
	1.20	2.6998	2.6998	2.7568	2.6953	0.0053	*
	1.30	2.7082	2.7082	2.7872	2.7061	0.0077	
	1.40	2.7124	2.7124	2.8159	2.7093	0.0082	
	1.50	2.7138	2.7138	2.8432	2.7090	0.0079	
	1.57	2.7138	2.7138	2.8613	2.7086	0.0076	
99%	0.00	2.9467	2.9467	2.9467	2.9381	0.0159	
	0.10	3.0405	3.0405	3.0405	3.0364	0.0098	*
	0.20	3.1208	3.1208	3.1208	3.1187	0.0186	
	0.30	3.1907	3.1907	3.1907	3.1902	0.0175	
	0.40	3.2529	3.2529	3.2529	3.2611	0.0145	
	0.50	3.3084	3.3084	3.3084	3.2960	0.0133	
	0.60	3.3585	3.3585	3.3587	3.3569	0.0112	*
	0.70	3.4039	3.4039	3.4048	3.3911	0.0153	
	0.80	3.4445	3.4443	3.4472	3.4244	0.0149	
	0.90	3.4796	3.4796	3.4862	3.4700	0.0140	
	1.00	3.5091	3.5089	3.5226	3.4641	0.0149	
	1.10	3.5320	3.5320	3.5568	3.5154	0.0116	*
	1.20	3.5486	3.5486	3.5888	3.5527	0.0112	*
	1.30	3.5591	3.5591	3.6189	3.5614	0.0117	*
	1.40	3.5644	3.5644	3.6475	3.5767	0.0106	*
	1.50	3.5661	3.5661	3.6746	3.5595	0.0148	
	1.57	3.5661	3.5661	3.6927	3.5592	0.0110	*

Table 5.6: Critical values for simple linear regression

d.f. = 20

cl	θ*	cv_ext	cv_nai	cv_app	cv_simu	s.e.	
	0.00	1.7247	1.7247	1.7247	1.7089	0.0050	
	0.10	1.8063	1.8063	1.8063	1.7968	0.0052	
	0.20	1.8788	1.8788	1.8788	1.8790	0.0036	*
	0.30	1.9436	1.9436	1.9436	1.9479	0.0036	*
	0.40	2.0021	2.0021	2.0021	1.9982	0.0036	*
	0.50	2.0549	2.0549	2.0553	2.0568	0.0054	
	0.60	2.1026	2.1026	2.1037	2.1032	0.0052	
	0.70	2.1447	2.1447	2.1483	2.1361	0.0053	
90%	0.80	2.1809	2.1807	2.1895	2.1810	0.0036	*
	0.90	2.2109	2.2109	2.2276	2.2071	0.0055	
	1.00	2.2343	2.2343	2.2633	2.2308	0.0053	
	1.10	2.2518	2.2518	2.2966	2.2488	0.0037	*
	1.20	2.2639	2.2637	2.3279	2.2627	0.0050	
	1.30	2.2709	2.2709	2.3573	2.2667	0.0055	
	1.40	2.2745	2.2743	2.3851	2.2747	0.0052	
	1.50	2.2755	2.2755	2.4116	2.2699	0.0051	
	1.57	2.2757	2.2757	2.4291	2.2675	0.0054	
	0.00	2.0860	2.0860	2.0860	2.0896	0.0049	*
	0.10	2.1691	2.1691	2.1691	2.1667	0.0048	*
	0.20	2.2419	2.2417	2.2419	2.2253	0.0069	
	0.30	2.3062	2.3062	2.3062	2.2981	0.0045	*
	0.40	2.3637	2.3637	2.3637	2.3619	0.0046	*
	0.50	2.4156	2.4156	2.4156	2.4111	0.0067	
	0.60	2.4623	2.4623	2.4629	2.4601	0.0077	
	0.70	2.5043	2.5043	2.5060	2.5206	0.0072	
95%	0.80	2.5410	2.5410	2.5458	2.5422	0.0078	
	0.90	2.5721	2.5721	2.5828	2.5706	0.0073	
	1.00	2.5973	2.5973	2.6169	2.5910	0.0072	
	1.10	2.6163	2.6163	2.6491	2.6059	0.0071	
	1.20	2.6295	2.6295	2.6791	2.6313	0.0065	
	1.30	2.6377	2.6377	2.7075	2.6363	0.0072	
	1.40	2.6417	2.6417	2.7342	2.6425	0.0049	*
	1.50	2.6428	2.6428	2.7595	2.6501	0.0050	*
	1.57	2.6430	2.6430	2.7765	2.6449	0.0069	
	0.00	2.8453	2.8453	2.8453	2.8386	0.0099	*
	0.10	2.9332	2.9332	2.9332	2.9225	0.0124	
	0.20	3.0079	3.0079	3.0079	2.9913	0.0152	
	0.30	3.0727	3.0727	3.0727	3.0766	0.0147	
	0.40	3.1299	3.1299	3.1299	3.1259	0.0104	*
	0.50	3.1810	3.1810	3.1810	3.1807	0.0141	
	0.60	3.2272	3.2270	3.2272	3.2252	0.0144	
	0.70	3.2687	3.2687	3.2691	3.2535	0.0096	*
99%	0.80	3.3061	3.3059	3.3076	3.3163	0.0095	*
	0.90	3.3387	3.3387	3.3433	3.3333	0.0090	*
	1.00	3.3661	3.3661	3.3762	3.3479	0.0128	
	1.10	3.3879	3.3879	3.4071	3.3824	0.0100	*
	1.20	3.4035	3.4035	3.4359	3.3848	0.0120	
	1.30	3.4134	3.4134	3.4632	3.4028	0.0101	*
	1.40	3.4184	3.4184	3.4887	3.4343	0.0129	
	1.50	3.4201	3.4201	3.5129	3.4159	0.0155	
	1.57	3.4203	3.4203	3.5291	3.4212	0.0102	*

Table 5.7: Critical values for simple linear regression

d.f. = 30

cl	θ	cv_ext	cv_nai	cv_app	cv_simu	s.e.	
	0.00	1.6973	1.6973	1.6973	1.6983	0.0037	*
	0.10	1.7764	1.7764	1.7764	1.7752	0.0048	
	0.20	1.8467	1.8467	1.8467	1.8450	0.0049	
	0.30	1.9095	1.9095	1.9095	1.9041	0.0049	
	0.40	1.9659	1.9657	1.9659	1.9601	0.0049	
	0.50	2.0168	2.0168	2.0168	2.0162	0.0047	
	0.60	2.0625	2.0625	2.0633	2.0614	0.0035	*
	0.70	2.1033	2.1033	2.1060	2.1053	0.0034	*
90%	0.80	2.1384	2.1384	2.1455	2.1268	0.0047	
	0.90	2.1676	2.1676	2.1819	2.1665	0.0049	
	1.00	2.1906	2.1906	2.2156	2.1866	0.0050	
	1.10	2.2078	2.2078	2.2473	2.2041	0.0051	
	1.20	2.2194	2.2194	2.2770	2.2183	0.0037	*
	1.30	2.2265	2.2265	2.3048	2.2208	0.0036	*
	1.40	2.2299	2.2299	2.3311	2.2285	0.0050	
	1.50	2.2309	2.2309	2.3561	2.2332	0.0048	
	1.57	2.2311	2.2311	2.3729	2.2239	0.0047	
	0.00	2.0423	2.0423	2.0423	2.0407	0.0046	*
	0.10	2.1222	2.1222	2.1222	2.1204	0.0063	
	0.20	2.1920	2.1918	2.1920	2.1931	0.0069	
	0.30	2.2534	2.2534	2.2534	2.2496	0.0063	
	0.40	2.3081	2.3081	2.3081	2.3010	0.0045	*
	0.50	2.3573	2.3573	2.3573	2.3552	0.0064	
	0.60	2.4017	2.4017	2.4021	2.4023	0.0045	*
	0.70	2.4417	2.4415	2.4429	2.4360	0.0069	
95%	0.80	2.4768	2.4768	2.4802	2.4795	0.0062	
	0.90	2.5065	2.5065	2.5149	2.4996	0.0046	*
	1.00	2.5307	2.5307	2.5471	2.5252	0.0045	*
	1.10	2.5492	2.5492	2.5771	2.5361	0.0068	
	1.20	2.5620	2.5620	2.6051	2.5609	0.0062	
	1.30	2.5700	2.5700	2.6316	2.5692	0.0046	*
	1.40	2.5738	2.5738	2.6566	2.5768	0.0068	
	1.50	2.5752	2.5752	2.6800	2.5656	0.0063	
	1.57	2.5752	2.5752	2.6958	2.5705	0.0066	
	0.00	2.7500	2.7500	2.7500	2.7273	0.0081	*
	0.10	2.8323	2.8321	2.8323	2.8278	0.0134	
	0.20	2.9019	2.9019	2.9019	2.8936	0.0120	
	0.30	2.9620	2.9620	2.9620	2.9523	0.0102	*
	0.40	3.0148	3.0148	3.0148	3.0035	0.0081	*
	0.50	3.0617	3.0617	3.0617	3.0600	0.0113	
	0.60	3.1040	3.1040	3.1040	3.1005	0.0146	
	0.70	3.1421	3.1421	3.1423	3.1382	0.0104	
99%	0.80	3.1764	3.1763	3.1774	3.1838	0.0098	*
	0.90	3.2066	3.2066	3.2096	3.2053	0.0137	
	1.00	3.2323	3.2321	3.2395	3.2264	0.0093	*
	1.10	3.2525	3.2525	3.2674	3.2540	0.0093	*
	1.20	3.2674	3.2674	3.2933	3.2507	0.0085	*
	1.30	3.2769	3.2769	3.3179	3.2657	0.0086	*
	1.40	3.2817	3.2817	3.3408	3.2828	0.0093	*
	1.50	3.2832	3.2832	3.3627	3.2830	0.0116	
	1.57	3.2834	3.2834	3.3772	3.2827	0.0166	

Table 5.8: Critical values for simple linear regression

d.f. = 40

cl	θ^*	cv_ext	cv_nai	cv_app	cv_simu	s.e.	
	0.00	1.6839	1.6837	1.6839	1.6826	0.0050	
	0.10	1.7619	1.7619	1.7619	1.7600	0.0035	*
	0.20	1.8311	1.8311	1.8311	1.8303	0.0047	
	0.30	1.8927	1.8927	1.8927	1.8848	0.0044	
	0.40	1.9482	1.9482	1.9482	1.9491	0.0051	
	0.50	1.9981	1.9981	1.9981	2.0005	0.0036	*
	0.60	2.0431	2.0431	2.0437	2.0454	0.0033	*
	0.70	2.0831	2.0831	2.0854	2.0816	0.0032	*
90%	0.80	2.1178	2.1178	2.1239	2.1178	0.0046	
	0.90	2.1466	2.1466	2.1596	2.1453	0.0034	*
	1.00	2.1693	2.1693	2.1926	2.1699	0.0033	*
	1.10	2.1863	2.1863	2.2234	2.1894	0.0034	*
	1.20	2.1977	2.1977	2.2522	2.1967	0.0048	
	1.30	2.2048	2.2046	2.2795	2.1997	0.0051	
	1.40	2.2080	2.2080	2.3050	2.2047	0.0047	
	1.50	2.2091	2.2091	2.3292	2.2141	0.0049	
	1.57	2.2093	2.2093	2.3454	2.2095	0.0033	*
	0.00	2.0212	2.0210	2.0212	2.0186	0.0064	
	0.10	2.0995	2.0995	2.0995	2.0973	0.0045	*
	0.20	2.1678	2.1676	2.1678	2.1665	0.0046	*
	0.30	2.2278	2.2276	2.2278	2.2294	0.0062	
	0.40	2.2812	2.2812	2.2812	2.2752	0.0044	*
	0.50	2.3291	2.3291	2.3292	2.3265	0.0041	*
	0.60	2.3723	2.3723	2.3725	2.3704	0.0041	*
	0.70	2.4112	2.4112	2.4122	2.4056	0.0060	
95%	0.80	2.4455	2.4455	2.4486	2.4385	0.0062	
	0.90	2.4749	2.4749	2.4821	2.4643	0.0055	
	1.00	2.4987	2.4985	2.5134	2.4939	0.0059	
	1.10	2.5168	2.5166	2.5424	2.5163	0.0062	
	1.20	2.5294	2.5294	2.5694	2.5230	0.0043	*
	1.30	2.5372	2.5370	2.5960	2.5366	0.0062	
	1.40	2.5410	2.5410	2.6190	2.5364	0.0061	
	1.50	2.5422	2.5422	2.6419	2.5422	0.0067	
	1.57	2.5424	2.5424	2.6570	2.5437	0.0046	*
	0.00	2.7044	2.7044	2.7044	2.7154	0.0082	*
	0.10	2.7841	2.7841	2.7841	2.7770	0.0123	
	0.20	2.8514	2.8514	2.8514	2.8459	0.0137	
	0.30	2.9092	2.9092	2.9092	2.8942	0.0129	
	0.40	2.9599	2.9599	2.9599	2.9500	0.0119	
	0.50	3.0649	3.0649	3.0649	3.0338	0.0122	
	0.60	3.0453	3.0453	3.0453	3.0468	0.0081	*
	0.70	3.0819	3.0817	3.0819	3.0753	0.0096	
99%	0.80	3.1147	3.1147	3.1154	3.1164	0.0108	
	0.90	3.1437	3.1437	3.1461	3.1427	0.0112	
	1.00	3.1684	3.1684	3.1745	3.1579	0.0094	*
	1.10	3.1883	3.1883	3.2010	3.1813	0.0105	
	1.20	3.2026	3.2026	3.2258	3.1946	0.0124	
	1.30	3.2119	3.2119	3.2489	3.2015	0.0137	
	1.40	3.2165	3.2165	3.2708	3.2098	0.0109	
	1.50	3.2182	3.2180	3.2914	3.2158	0.0090	*
	1.57	3.2182	3.2182	3.3051	3.2170	0.0138	

Table 5.9: Critical values for simple linear regression

d.f. = 60

cl	θ	cv_ext	cv_nai	cv_app	cv_simu	s.e.	
	0.00	1.6706	1.6706	1.6706	1.6652	0.0034	*
	0.10	1.7476	1.7476	1.7476	1.7516	0.0046	
	0.20	1.8157	1.8157	1.8157	1.8133	0.0035	*
	0.30	1.8763	1.8763	1.8763	1.8713	0.0046	
	0.40	1.9306	1.9306	1.9306	1.9281	0.0034	*
	0.50	1.9798	1.9798	1.9798	1.9856	0.0050	
	0.60	2.0240	2.0240	2.0244	2.0237	0.0045	
	0.70	2.0633	2.0633	2.0654	2.0608	0.0045	
90%	0.80	2.0974	2.0974	2.1030	2.0955	0.0034	*
	0.90	2.1258	2.1258	2.1377	2.1232	0.0045	
	1.00	2.1483	2.1483	2.1699	2.1463	0.0049	
	1.10	2.1649	2.1649	2.2000	2.1642	0.0034	*
	1.20	2.1764	2.1764	2.2280	2.1743	0.0044	
	1.30	2.1832	2.1832	2.2546	2.1826	0.0033	*
	1.40	2.1866	2.1866	2.2795	2.1846	0.0045	
	1.50	2.1878	2.1878	2.3029	2.1863	0.0035	*
	1.57	2.1878	2.1878	2.3186	2.1904	0.0047	
	0.00	2.0004	2.0002	2.0004	1.9950	0.0066	
	0.10	2.0772	2.0772	2.0772	2.0780	0.0042	*
	0.20	2.1439	2.1439	2.1439	2.1384	0.0062	
	0.30	2.2027	2.2027	2.2027	2.1996	0.0046	*
	0.40	2.2549	2.2547	2.2549	2.2589	0.0068	
	0.50	2.3016	2.3016	2.3016	2.3008	0.0042	*
	0.60	2.3437	2.3437	2.3437	2.3415	0.0047	*
	0.70	2.3817	2.3817	2.3822	2.3841	0.0044	*
95%	0.80	2.4150	2.4150	2.4177	2.4180	0.0043	*
	0.90	2.4438	2.4438	2.4501	2.4395	0.0056	
	1.00	2.4671	2.4671	2.4804	2.4702	0.0041	*
	1.10	2.4850	2.4850	2.5084	2.4791	0.0041	*
	1.20	2.4974	2.4974	2.5348	2.5028	0.0064	
	1.30	2.5050	2.5050	2.5593	2.5056	0.0043	*
	1.40	2.5088	2.5088	2.5826	2.5078	0.0058	
	1.50	2.5100	2.5100	2.6045	2.5094	0.0061	
	1.57	2.5102	2.5102	2.6190	2.5078	0.0042	*
	0.00	2.6602	2.6602	2.6602	2.6544	0.0079	*
	0.10	2.7374	2.7374	2.7374	2.7313	0.0079	*
	0.20	2.8024	2.8024	2.8024	2.7978	0.0136	
	0.30	2.8581	2.8581	2.8581	2.8581	0.0128	
	0.40	2.9069	2.9067	2.9069	2.9066	0.0131	
	0.50	2.9500	2.9500	2.9500	2.9479	0.0077	*
	0.60	2.9887	2.9885	2.9887	2.9881	0.0082	*
	0.70	3.0236	3.0236	3.0236	3.0153	0.0080	*
99%	0.80	3.0550	3.0550	3.0564	3.0461	0.0119	
	0.90	3.0828	3.0828	3.0847	3.0799	0.0081	*
	1.00	3.1067	3.1067	3.1118	3.0903	0.0102	
	1.10	3.1259	3.1259	3.1370	3.1044	0.0124	
	1.20	3.1398	3.1398	3.1604	3.1382	0.0101	
	1.30	3.1488	3.1488	3.1824	3.1477	0.0089	*
	1.40	3.1536	3.1536	3.2031	3.1445	0.0102	
	1.50	3.1551	3.1549	3.2226	3.1599	0.0075	*
	1.57	3.1551	3.1551	3.2355	3.1428	0.0106	

Table 5.10: Critical values for simple linear regression

5.5.2 For polynomial regression of various orders

In this subsection, we compare the conservative method of Naiman (1986), the approximate method of Sun and Loader (1994), and the simulation-based method of Liu, Wynn and Hayter (2005) for an one-dimensional polynomial regression of $(p-1)$th order. In our comparison, we set $p = 3, 4, 5$ respectively corresponding to the quadratic regression, the cubic regression, and the 4th order polynomial regression.

For Naiman's method, we calculate the critical values via the following formula

$$\begin{aligned} 1-\alpha &= 1-\int_0^{1/c} \min\{F_{p-2,2}[2((ct)^{-2}-1)/(p-2)] \times \Lambda(\gamma)/\pi \\ &\quad + F_{p-1,1}[((ct)^{-2}-1)/(p-1)], 1\} f_T(t)dt, \end{aligned} \tag{5.34}$$

where f_T is the density function of the random variable T such that $pT^2 \sim F_{\nu,p}$, the F distribution with $\nu = n-p$ and p degrees of freedom, c is a critical value, and $\Lambda(\gamma)$ can be obtained from

$$\Lambda(\gamma) = \int_a^b \|\gamma'(\mathrm{x})\| dx,$$

where $\gamma'(\mathrm{x})$ denotes the derivative of $\gamma(\mathrm{x})$ with $\mathrm{x} = (1, x, x^2, \ldots, x^{p-1})$ for all $x \in [a, b]$. Specifically, $\Lambda(\gamma)$ can be calculated in the following way. We have

$$\begin{aligned} \gamma'(\mathrm{x}) &= \left(\frac{P\mathrm{x}}{\|P\mathrm{x}\|}\right)' \\ &= \frac{(P\mathrm{x})'\|P\mathrm{x}\| - (\|P\mathrm{x}\|)'(P\mathrm{x})}{\|P\mathrm{x}\|^2} \\ &= \frac{(P\mathrm{x})'(\|P\mathrm{x}\|^2)^{1/2} - [(\|P\mathrm{x}\|^2)^{1/2}]'(P\mathrm{x})}{\|P\mathrm{x}\|^2} \\ &= \frac{(P\mathrm{x})'(\|P\mathrm{x}\|^2)^{1/2} - \frac{1}{2}(\|P\mathrm{x}\|^2)^{-1/2}(\|P\mathrm{x}\|^2)'(P\mathrm{x})}{\|P\mathrm{x}\|^2}, \end{aligned}$$

where $P\mathrm{x} = (\mathrm{p}_0, \mathrm{p}_1, \ldots, \mathrm{p}_{p-1})(1, x, x^2, \ldots, x^{p-1})^T = \mathrm{p}_0 + \mathrm{p}_1 x + \mathrm{p}_2 x^2 + \cdots + \mathrm{p}_{p-1} x^{p-1}$, $\mathrm{p}_0, \ldots, \mathrm{p}_{\mathrm{p}-1}$ are the p columns of the matrix P. Note that $\|P\mathrm{x}\|^2$

is the polynomial of order $2p-2$, whose coefficients can be obtained by using commands conv and sum, $(P\mathbf{x})'$ and $(\|P\mathbf{x}\|^2)'$ are the polynomials of order $p-2$ and $2p-3$ respectively, where their coefficients can be obtained by using command polyder. Then, $\gamma'(\mathbf{x})$ can be found. By using command quad to implement the numerical integration, we can easily compute $\Lambda(\gamma)$.

For the approximate method, critical values can be calculated from

$$\alpha = \frac{\kappa_0}{\pi}(1+\frac{c^2}{\nu})^{-\nu/2} + \mathrm{P}\{|t_\nu| > c\}, \tag{5.35}$$

where κ_0 is the length of the path on the surface S^{n-1} of the unit sphere. Note that $\kappa_0 = \Lambda(\gamma)$ from (5.32).

For the simulation-based method, we obtain the critical values by following the procedure in Section 5.3.

From (5.34) and (5.35), it is clear that both Naiman's method and the approximate method depend on the length of the path, the degree of freedom and the confidence level. However, it is not clear that the simulation-based method depends on the same factors. So we use three general common factors here, i.e., the design matrix, the restricted interval for x and the preassigned confidence level.

First, we come to choose the design matrix. This can be done by choosing some design points of different locations on preassigned design intervals. Now we choose three design intervals [-1, 1], [0, 2] and [-2, 0]. For each interval, we have four structures of 8 design points. Take the case when the interval considered is [-1, 1] for example, the four structures are:

1. S_1=[-0.2 -0.16 -0.13 -0.06 0 0.07 0.11 0.18], where the design points are distributed around the middle of the interval,

2. S_2=[-1 -0.95 -0.9 -0.89 0.92 0.95 0.98 1], where the design points concentrate on the two ends of the interval,

3. S_3=[0.86 0.89 0.91 0.93 0.96 0.98 0.99 1], where the design points are near the upper bound of the interval,

4. S_4=[-1 -0.7 -0.4 -0.1 0.2 0.5 0.8 1.1], where all the design points are equally spaced.

Such four structures of 8 design points are also adopted for the design intervals [0, 2] and [-2, 0]. So we have another 8 sets (four structures $\times$ two design intervals) of 8 design points given by

$$\begin{aligned}
S_5 &= [0.85\ 0.88\ 0.93\ 0.97\ 1\ 1.04\ 1.09\ 1.12], \\
S_6 &= [-0.05\ -0.02\ 0\ 0.03\ 1.94\ 1.99\ 2\ 2.03], \\
S_7 &= [-0.01\ -0.03\ 0\ 0.02\ 0.05\ 0.06\ 0.09\ 0.1], \\
S_8 &= [0\ 0.3\ 0.6\ 0.9\ 1.2\ 1.5\ 1.8\ 2.1], \\
S_9 &= [-1.11\ -1.08\ -1.01\ -1\ -0.99\ -0.97\ -0.95\ -0.92], \\
S_{10} &= [-2\ -1.99\ -1.96\ -1.91\ -0.08\ -0.05\ -0.01\ 0.02], \\
S_{11} &= [-0.1\ -0.07\ -0.02\ 0\ 0.03\ 0.09\ 0.15], \\
S_{12} &= [-2\ -1.7\ -1.4\ -1.1\ -0.8\ -0.5\ -0.2\ 0.1].
\end{aligned}$$

Therefore, 12 design matrices can be obtained so far which are marked by $D1, \ldots, D12$. In addition, it is motivated to choose extra 12 design matrices of 35 design points in order to make the degree of freedom be both small ($\nu < 5$) and large ($\nu > 30$) in our comparison. We choose these extra 12 design matrices also of the same four structures as described previously, still on the three design intervals [-1, 1], [0, 2] and [-2, 0] respectively. The 12 sets (four structures $\times$ three design intervals) of 35 design points are:

1. S_{13}=[-0.39 -0.36 -0.31 -0.29 -0.23 -0.21 -0.2 -0.18 -0.17 -0.14 -0.13 -0.11 -0.1 -0.07 -0.06 -0.03 -0.01 0 0.02 0.03 0.05 0.09 0.11 0.13 0.14 0.18 0.2 0.21 0.24 0.25 0.3 0.32 0.36 0.39 0.4],

2. S_{14}=[-1.23 -1.21 -1.15 -1.1 -1.08 -1.02 -1 -0.99 -0.95 -0.94 -0.92 -0.89 -0.82 -0.73 -0.72 -0.71 -0.7 0.71 0.73 0.76 0.79 0.81 0.84 0.85 0.88 0.92 0.95 0.98 0.99 1.02 1.04 1.08 1.12 1.13 1.19],

3. S_{15}=[0.71 0.72 0.74 0.75 0.77 0.78 0.81 0.82 0.85 0.88 0.9 0.91 0.93 0.94

0.98 1 1.01 1.07 1.08 1.09 1.11 1.12 1.13 1.16 1.19 1.2 1.22 1.25 1.26 1.28 1.29 1.3 1.31 1.35 1.4],

4. S_{16}=[-1.36 -1.28 -1.2 -1.12 -1.04 -0.96 -0.88 -0.8 -0.72 -0.64 -0.56 -0.48 -0.4 -0.32 -0.24 -0.16 -0.08 0 0.08 0.16 0.24 0.32 0.4 0.48 0.56 0.64 0.72 0.8 0.88 0.96 1.04 1.12 1.2 1.28 1.36],

5. S_{17}=[0.65 0.69 0.71 0.72 0.75 0.76 0.78 0.82 0.85 0.88 0.89 0.9 0.93 0.95 0.97 0.99 1 1.02 1.04 1.07 1.08 1.13 1.15 1.19 1.2 1.21 1.23 1.25 1.26 1.28 1.3 1.32 1.33 1.35 1.4],

6. S_{18}=[-0.13 -0.12 -0.1 -0.09 -0.07 -0.05 -0.03 -0.01 0 0.03 0.05 0.09 0.12 0.16 0.17 0.18 0.23 0.28 0.31 0.35 1.72 1.76 1.81 1.83 1.85 1.94 1.95 1.97 1.99 2 2.02 2.05 2.07 2.14 2.16],

7. S_{19}=[1.68 1.71 1.73 1.76 1.77 1.79 1.8 1.82 1.85 1.87 1.88 1.91 1.92 1.94 1.95 1.99 2 2.01 2.03 2.07 2.12 2.13 2.15 2.18 2.19 2.24 2.26 2.29 2.31 2.33 2.34 2.37 2.39 2.4 2.42],

8. S_{20}=[-0.36 -0.28 -0.2 -0.12 -0.04 0.04 0.12 0.2 0.28 0.36 0.44 0.52 0.6 0.68 0.76 0.84 0.92 1 1.08 1.16 1.24 1.32 1.4 1.48 1.56 1.64 1.72 1.8 1.88 1.96 2.04 2.12 2.2 2.28 2.36],

9. S_{21}=[-1.7 -1.62 -1.5 -1.43 -1.41 -1.38 -1.37 -1.35 -1.31 -1.28 -1.22 -1.2 -1.18 -1.15 -1.13 -1.11 -1.09 -1.07 -1.05 -1.04 -1.02 -1.01 -0.98 -0.96 -0.95 -0.92 -0.85 -0.81 -0.78 -0.72 -0.71 -0.69 -0.68 -0.65 -0.61],

10. S_{22}=[-2.36 -2.31 -2.28 -2.27 -2.25 -2.2 -2.12 -2.09 -2.06 -2.01 -2 -1.99 -1.97 -1.93 -1.91 -1.86 -1.81 -0.53 -0.44 -0.42 -0.41 -0.38 -0.35 -0.33 -0.31 -0.27 -0.26 -0.24 -0.2 -0.15 -0.12 -0.11 -0.08 -0.05 0],

11. S_{23}=[-2.36 -2.31 -2.29 -2.28 -2.26 -2.25 -2.2 -2.18 -2.12 -2.09 -2.06 -2.01 -2 -1.99 -1.97 -1.93 -1.91 -1.86 -1.85 -1.81 -1.77 -1.75 -1.73 -1.72 -1.68 -1.66 -1.64 -1.61 -1.56 -1.53 -1.52 -1.49 -1.48 -1.45 -1.41],

12. S_{24}=[-2.2 -2.13 -2.06 -1.99 -1.92 -1.85 -1.78 -1.71 -1.64 -1.57 -1.5 -1.43 -1.36 -1.29 -1.22 -1.15 -1.08 -1.01 -0.94 -0.87 -0.8 -0.73 -0.66 -0.59 -0.52 -0.45 -0.38 -0.31 -0.24 -0.17 -0.1 -0.03 0.04 0.11 0.18].

We denote the 12 design matrices obtained from $S_{13}, \ldots, S_{24}$ by $D13, \ldots, D24$. Consequently, we have totally 24 design matrices $D1, \ldots, D24$ for our comparison.

Next, we choose the restricted intervals of x on which confidence bands are constructed. For each design interval, we choose two restricted intervals. One is the same as the design interval, the other is of smaller length. Specifically, these six restricted intervals are: [-1, 1], [-0.7, 0.7], [0, 2], [0.6, 1.4], [-2, 0] and [-1.5, -0.5]. Note that since both the design matrices and the restricted intervals are chosen according to the design intervals, in our comparison, the first two restricted intervals are used together with $D1, \ldots, D4$ and $D13, \ldots, D16$, the middle two with $D5, \ldots, D8$ and $D17, \ldots, D20$, and the last two with $D9, \ldots, D12$ and $D21, \ldots, D24$.

Finally, 90% and 95% confidence levels are employed. The choice of such designs is to obtain as many combinations of the three factors as possible such that our comparison gives a general view.

We calculate the critical values of the confidence bands based on the three methods for polynomial regression of up to the 4th order, and record them in Tables 5.11-5.16. The columns titled κ_0 contain the values of the length of the path.

From the results, we may draw some conclusions. When the degree of freedom is small ($\nu < 5$), Naiman's method and the approximate method of Sun and Loader (1994) have almost the same critical values for $\kappa_0 < 0.6$; when $\kappa_0 > 0.6$, the critical values of Naiman's method are generally smaller than those of the approximate method and the difference between the critical values of these two methods follows that:

$0.6 < \kappa_0 < 1.5$	the difference is at the third decimal place,
$1.5 < \kappa_0 < 3$	the difference is at the second decimal place,
$3 < \kappa_0 < 4$	the difference is around 0.15,
$4 < \kappa_0 < 5$	the difference is around 0.25,
$5 < \kappa_0 < 6$	the difference is around 0.3,
$6 < \kappa_0 < 7$	the difference is around 0.5,
$7 < \kappa_0 < 8.5$	the difference is around 0.7,
$8.5 < \kappa_0 < 10$	the difference is around 0.9.

The critical values of the simulation-based method are even smaller than those of Naiman's method, with the difference generally at the second decimal place. When $\nu > 30$, the distinctions among the critical values of the three methods are not evident relative to the case when $\nu < 5$, generally at the second decimal place. Also, the simulation-based method obtains the smallest critical values.

Consider that for a large number of simulations, the simulation-based method seems to be able to compute as accurate critical values as the exact method. Therefore, we may conclude that Naiman's method is good enough because it is basically a conservative method but its critical values are not much conservative actually, the approximate method is not good as its critical values are even larger than those of the conservative method. However, in particular, three methods give almost the same critical values when $\kappa_0 < 1$.

d.f. = 5

RI	DM	k0	cl	cv nai	cv app	cv simu	s.e.
(-1, 1)	D1	5.3026	90%	3.1396	3.4360	3.0844	0.0106
			95%	3.8404	4.1381	3.7748	0.0171
	D2	4.5157	90%	3.0947	3.3124	3.0631	0.0152
			95%	3.7872	4.0018	3.7639	0.0221
	D3	4.5376	90%	3.0961	3.3160	3.0728	0.0114
			95%	3.7890	4.0057	3.7385	0.0206
	D4	3.5219	90%	3.0091	3.1334	3.0200	0.0119
			95%	3.6863	3.8048	3.7209	0.0176
(-0.7, 0.7)	D1	5.2056	90%	3.1348	3.4215	3.1091	0.0141
			95%	3.8348	4.1220	3.7943	0.0185
	D2	0.3200	90%	2.1823	2.1824	2.1754	0.0119
			95%	2.7583	2.7584	2.7590	0.0173
	D3	0.1544	90%	2.0990	2.0989	2.1068	0.0095
			95%	2.6651	2.6651	2.6728	0.0192
	D4	2.2914	90%	2.8244	2.8606	2.8286	0.0116
			95%	3.4734	3.5053	3.5166	0.0200
(0, 2)	D5	5.3104	90%	3.1400	3.4372	3.1004	0.0115
			95%	3.8408	4.1394	3.7724	0.0195
	D6	3.0409	90%	2.9499	3.0349	2.9222	0.0128
			95%	3.6176	3.6966	3.6108	0.0186
	D7	3.5613	90%	3.0132	3.1411	2.9373	0.0115
			95%	3.6912	3.8133	3.6221	0.0188
	D8	3.5219	90%	3.0091	3.1334	3.0330	0.0135
			95%	3.6863	3.8048	3.7318	0.0183
(0.6, 1.4)	D5	5.0487	90%	3.1267	3.3977	3.0955	0.0134
			95%	3.8251	4.0957	3.7755	0.0191
	D6	0.0637	90%	2.0504	2.0504	2.0562	0.0113
			95%	2.6105	2.6106	2.5969	0.0167
	D7	0.0373	90%	2.0359	2.0359	2.0523	0.0114
			95%	2.5943	2.5942	2.5970	0.0144
	D8	0.9811	90%	2.4608	2.4619	2.5057	0.0113
			95%	3.0668	3.0677	3.1145	0.0143
(-2, 0)	D9	5.3279	90%	3.1408	3.4398	3.0870	0.0123
			95%	3.8418	4.1422	3.7878	0.0182
	D10	4.1462	90%	3.0676	3.2492	3.0676	0.0135
			95%	3.7549	3.9321	3.7248	0.0181
	D11	2.5200	90%	3.0811	3.1579	3.0077	0.0145
			95%	3.8793	3.9553	3.7830	0.0245
	D12	3.5219	90%	3.0091	3.1334	3.0230	0.0142
			95%	3.6863	3.8048	3.6984	0.0225
(-1.5, -0.5)	D9	5.2084	90%	3.1350	3.4219	3.0795	0.0128
			95%	3.8350	4.1225	3.7704	0.0183
	D10	0.1205	90%	2.0811	2.0811	2.0536	0.0105
			95%	2.6451	2.6451	2.6566	0.0197
	D11	0.0982	90%	2.1914	2.1914	2.1994	0.0123
			95%	2.8460	2.8460	2.8239	0.0184
	D12	1.3623	90%	2.5893	2.5944	2.6163	0.0125
			95%	3.2094	3.2134	3.2979	0.0181

Table 5.11: Critical values for quadratic regression

d.f. = 4

RI	DM	k0	cl	cv_nai	cv_app	cv_simu	s.e.
(-1, 1)	D1	7.4228	90%	3.6452	4.1269	3.5498	0.0158
			95%	4.5509	5.0693	4.4359	0.0349
	D2	6.0073	90%	3.5609	3.9054	3.4880	0.0157
			95%	4.4482	4.8138	4.3232	0.0271
	D3	6.5732	90%	3.5982	3.9978	3.1112	0.0153
			95%	4.4936	4.9202	3.9330	0.0289
	D4	4.9461	90%	3.4728	3.7157	3.4598	0.0173
			95%	4.3415	4.5952	4.3415	0.0244
(-0.7, 0.7)	D1	7.3262	90%	3.6403	4.1128	3.5432	0.0156
			95%	4.5450	5.0529	4.4449	0.0283
	D2	1.2913	90%	2.7504	2.7557	2.7376	0.0137
			95%	3.4909	3.4955	3.4896	0.0206
	D3	0.1528	90%	2.2233	2.2234	2.2277	0.0118
			95%	2.8833	2.8832	2.8176	0.0166
	D4	2.7636	90%	3.1614	3.2249	3.0935	0.0150
			95%	3.9697	4.0322	3.9233	0.0228
(0, 2)	D5	7.6387	90%	3.6555	4.1580	3.3457	0.0151
			95%	4.5636	5.1052	4.1663	0.0304
	D6	4.1500	90%	3.3852	3.5555	3.2117	0.0155
			95%	4.2361	4.4111	4.0381	0.0204
	D7	4.9113	90%	3.4693	3.7090	3.2998	0.0154
			95%	4.3375	4.5875	4.1284	0.0234
	D8	4.9461	90%	3.4728	3.7157	3.3164	0.0132
			95%	4.3415	4.5952	4.1728	0.0251
(0.6, 1.4)	D5	7.4255	90%	3.6452	4.1273	3.3499	0.0144
			95%	4.5511	5.0697	4.2144	0.0253
	D6	0.7692	90%	2.5377	2.5385	2.5121	0.0152
			95%	3.2461	3.2467	3.2211	0.0209
	D7	0.0367	90%	2.1545	2.1545	2.1529	0.0117
			95%	2.8028	2.8028	2.8375	0.0234
	D8	1.6600	90%	2.8765	2.8898	2.7966	0.0113
			95%	3.6367	3.6488	3.5487	0.0219
(-2, 0)	D9	7.7237	90%	3.6597	4.1702	3.2333	0.0139
			95%	4.5686	5.1192	4.0554	0.0255
	D10	5.6165	90%	3.5315	3.8384	3.3497	0.0153
			95%	4.4125	4.7363	4.1687	0.0259
	D11	3.4267	90%	3.7321	3.9436	3.5419	0.0176
			95%	4.8915	5.1318	4.6957	0.0460
	D12	4.9461	90%	3.4728	3.7157	3.4284	0.0145
			95%	4.3415	4.5952	4.2536	0.0243
(-1.5, -0.5)	D9	7.6347	90%	3.6553	4.1575	3.2417	0.0157
			95%	4.5634	5.1046	4.0763	0.0254
	D10	0.9690	90%	2.6242	2.6259	2.6024	0.0119
			95%	3.3455	3.3470	3.3339	0.0203
	D11	0.0788	90%	2.4100	2.4101	2.4075	0.0144
			95%	3.2526	3.2526	3.2787	0.0247
	D12	1.9718	90%	2.9700	2.9933	2.6802	0.0137
			95%	3.7452	3.7671	3.6658	0.0232

Table 5.12: Critical values for cubic regression

d.f. = 3

RI	DM	k0	cl	cv_nai	cv_app	cv_simu	s.e.
(-1, 1)	D1	9.5821	90%	4.4008	5.3870	4.2972	0.0232
			95%	5.7397	6.9125	5.6521	0.0376
	D2	8.1877	90%	4.3208	5.1236	4.2971	0.0208
			95%	5.6370	6.5862	5.5020	0.0368
	D3	8.5104	90%	4.3411	5.1869	3.0442	0.0157
			95%	5.6632	6.6646	4.0336	0.0338
	D4	6.2884	90%	4.1716	4.7173	4.1753	0.0235
			95%	5.4458	6.0843	5.3716	0.0456
(-0.7, 0.7)	D1	9.4935	90%	4.3962	5.3710	4.2923	0.0217
			95%	5.7340	6.8926	5.5911	0.0489
	D2	0.2424	90%	2.5223	2.5223	2.5073	0.0157
			95%	3.3911	3.3911	3.3663	0.0205
	D3	0.1246	90%	2.4421	2.4421	2.4672	0.0167
			95%	3.2923	3.2923	3.2890	0.0250
	D4	3.7802	90%	3.8356	4.0542	3.7480	0.0193
			95%	5.0191	5.2678	4.9218	0.0311
(0, 2)	D5	9.6723	90%	4.4053	5.4033	3.1014	0.0190
			95%	5.7455	6.9326	4.1268	0.0342
	D6	5.2263	90%	4.0558	4.4585	3.8969	0.0185
			95%	5.2984	5.7651	5.1019	0.0352
	D7	6.4701	90%	4.1885	4.7591	4.0562	0.0205
			95%	5.4676	6.1358	5.2982	0.0472
	D8	6.2884	90%	4.1715	4.7173	4.1380	0.0268
			95%	5.4458	6.0843	5.4158	0.0488
(0.6, 1.4)	D5	9.4448	90%	4.3936	5.3621	3.1369	0.0180
			95%	5.7306	6.8816	4.0577	0.0310
	D6	0.0600	90%	2.3967	2.3967	2.3913	0.0147
			95%	3.2362	3.2360	3.2062	0.0249
	D7	0.0281	90%	2.3737	2.3738	2.3895	0.0157
			95%	3.2078	3.2077	3.2094	0.0247
	D8	2.1419	90%	3.4282	3.4852	3.4198	0.0195
			95%	4.5077	4.5702	4.5244	0.0323
(-2, 0)	D9	9.6721	90%	4.4053	5.4033	2.8066	0.0162
			95%	5.7455	6.9326	3.7636	0.0237
	D10	7.5293	90%	4.2754	4.9897	4.1819	0.0212
			95%	5.5789	6.4206	6.4942	0.0320
	D11	4.5032	90%	5.2095	6.0663	4.8847	0.0382
			95%	7.5187	7.9999	7.1059	0.0641
	D12	6.2884	90%	4.1715	4.7173	4.1431	0.0218
			95%	5.4458	6.0843	5.4060	0.0419
(-1.5, -0.5)	D9	9.5552	90%	4.3994	5.3822	2.8201	0.0181
			95%	5.7380	6.9065	3.7483	0.0275
	D10	0.0920	90%	2.4193	2.4194	2.4388	0.0142
			95%	3.2642	3.2641	3.2449	0.0208
	D11	0.0757	90%	2.9974	2.9973	3.0243	0.0215
			95%	4.4105	4.4105	4.3834	0.0411
	D12	2.7749	90%	3.6135	3.7234	3.5764	0.0176
			95%	4.7392	4.8619	4.6602	0.0326

Table 5.13: Critical values for 4th order polynomial regression

d.f. = 32

RI	DM	k0	cl	cv_nai	cv_app	cv_simu	s.e.
(-1, 1)	D13	4.9197	90%	2.4691	2.5355	2.4382	0.0065
			95%	2.8000	2.8462	2.7746	0.0089
	D14	3.1530	90%	2.3533	2.3675	2.3498	0.0064
			95%	2.6800	2.6875	2.6718	0.0082
	D15	2.5281	90%	2.2842	2.2887	2.2341	0.0071
			95%	2.6111	2.6131	2.5730	0.0086
	D16	2.9628	90%	2.3344	2.3449	2.3505	0.0067
			95%	2.6607	2.6662	2.6815	0.0083
(-0.7, 0.7)	D13	4.6777	90%	2.4582	2.5159	2.4377	0.0074
			95%	2.7883	2.8277	2.7728	0.0085
	D14	1.0013	90%	2.0125	2.0125	2.0029	0.0076
			95%	2.3505	2.3506	2.3554	0.0088
	D15	0.6914	90%	1.9315	1.9316	1.9275	0.0069
			95%	2.2725	2.2725	2.2476	0.0078
	D16	1.7779	90%	2.1720	2.1723	2.2024	0.0073
			95%	2.5028	2.5030	2.5401	0.0086
(0, 2)	D17	5.0496	90%	2.4747	2.5457	2.4522	0.0066
			95%	2.8057	2.8558	2.7696	0.0094
	D18	3.2945	90%	2.3664	2.3835	2.3757	0.0063
			95%	2.6931	2.7026	2.6776	0.0077
	D19	2.4902	90%	2.2793	2.2835	2.2359	0.0071
			95%	2.6054	2.6082	2.5724	0.0086
	D20	2.9628	90%	2.3344	2.3449	2.3519	0.0066
			95%	2.6607	2.6662	2.6945	0.0102
(0.6, 1.4)	D17	4.1075	90%	2.4273	2.4660	2.4333	0.0061
			95%	2.7557	2.7806	2.7461	0.0078
	D18	0.2998	90%	1.8088	1.8087	1.8073	0.0073
			95%	2.1521	2.1522	2.1339	0.0097
	D19	0.1976	90%	1.7719	1.7720	1.7763	0.0065
			95%	2.1156	2.1156	2.1129	0.0086
	D20	0.7848	90%	1.9573	1.9573	1.9816	0.0070
			95%	2.2973	2.2974	2.3296	0.0086
(-2, 0)	D21	4.6422	90%	2.4564	2.5130	2.4466	0.0064
			95%	2.7865	2.8249	2.7731	0.0074
	D22	3.1560	90%	2.3537	2.3678	2.3298	0.0067
			95%	2.6802	2.6878	2.6763	0.0082
	D23	2.9052	90%	2.3283	2.3379	2.2900	0.0077
			95%	2.6548	2.6595	2.6175	0.0080
	D24	3.4094	90%	2.3765	2.3962	2.3978	0.0071
			95%	2.7034	2.7146	2.7133	0.0079
(-1.5, -0.5)	D21	3.5553	90%	2.3884	2.4117	2.3952	0.0066
			95%	2.7157	2.7292	2.7239	0.0080
	D22	0.6325	90%	1.9147	1.9147	1.9245	0.0070
			95%	2.2561	2.2562	2.2638	0.0080
	D23	0.8735	90%	1.9805	1.9806	1.9765	0.0070
			95%	2.3199	2.3199	2.3127	0.0088
	D24	1.2993	90%	2.0799	2.0799	2.1136	0.0069
			95%	2.4152	2.4151	2.4545	0.0083

Table 5.14: Critical values for quadratic regression

d.f. = 31

RI	DM	k0	cl	cv_nai	cv_app	cv_simu	s.e.
(-1, 1)	D13	7.0573	90%	2.6222	2.6824	2.5529	0.0067
			95%	2.9472	2.9871	2.8868	0.0083
	D14	3.9602	90%	2.4463	2.4559	2.4316	0.0068
			95%	2.7680	2.7726	2.7421	0.0093
	D15	3.3519	90%	2.3890	2.3933	2.2818	0.0068
			95%	2.7117	2.7135	2.6065	0.0087
	D16	3.6551	90%	2.4189	2.4256	2.4021	0.0058
			95%	2.7411	2.7440	2.7063	0.0078
(-0.7, 0.7)	D13	6.7986	90%	2.6123	2.6675	2.5467	0.0064
			95%	2.9369	2.9729	2.8835	0.0085
	D14	1.5989	90%	2.1424	2.1424	2.1347	0.0065
			95%	2.4759	2.4760	2.4677	0.0084
	D15	0.8327	90%	1.9722	1.9721	1.9681	0.0076
			95%	2.3130	2.3130	2.2959	0.0083
	D16	2.2531	90%	2.2519	2.2522	2.2280	0.0074
			95%	2.5800	2.5801	2.5531	0.0094
(0, 2)	D17	7.1639	90%	2.6262	2.6884	2.4989	0.0072
			95%	2.9511	2.9927	2.8375	0.0088
	D18	4.2371	90%	2.4691	2.4816	2.3972	0.0065
			95%	2.7907	2.7971	2.7251	0.0095
	D19	3.3726	90%	2.3912	2.3955	2.2941	0.0073
			95%	2.7137	2.7157	2.6332	0.0106
	D20	3.6551	90%	2.4189	2.4256	2.3320	0.0065
			95%	2.7411	2.7440	2.6387	0.0086
(0.6, 1.4)	D17	6.0044	90%	2.5780	2.6180	2.4901	0.0063
			95%	2.9009	2.9260	2.8361	0.0090
	D18	0.9009	90%	1.9899	1.9898	1.9874	0.0068
			95%	2.3301	2.3300	2.3184	0.0090
	D19	0.2084	90%	1.7776	1.7777	1.7852	0.0068
			95%	2.1224	2.1223	2.1200	0.0093
	D20	1.4320	90%	2.1099	2.1098	2.0644	0.0064
			95%	2.4449	2.4450	2.3894	0.0082
(-2, 0)	D21	6.8471	90%	2.6143	2.6704	2.3760	0.0075
			95%	2.9388	2.9756	2.7088	0.0085
	D22	4.0813	90%	2.4564	2.4673	2.3510	0.0062
			95%	2.7782	2.7835	2.6671	0.0089
	D23	4.3276	90%	2.4761	2.4899	2.3165	0.0073
			95%	2.7976	2.8047	2.6534	0.0077
	D24	4.6027	90%	2.4963	2.5136	2.4314	0.0068
			95%	2.8178	2.8272	2.7686	0.0089
(-1.5, -0.5)	D21	4.8340	90%	2.5120	2.5327	2.3435	0.0063
			95%	2.8337	2.8452	2.6628	0.0092
	D22	1.1504	90%	2.0498	2.0497	2.0327	0.0073
			95%	2.3876	2.3875	2.3488	0.0094
	D23	1.4132	90%	2.1061	2.1061	2.0828	0.0071
			95%	2.4413	2.4414	2.4096	0.0091
	D24	1.8895	90%	2.1942	2.1943	2.1269	0.0062
			95%	2.5252	2.5252	2.4756	0.0090

Table 5.15: Critical values for cubic regression

d.f. = 30

RI	DM	k0	cl	cv_nai	cv_app	cv_simu	s.e.
(-1, 1)	D13	9.2648	90%	2.7391	2.7977	2.6963	0.0061
			95%	3.0604	3.0986	3.0410	0.0091
	D14	5.0910	90%	2.5463	2.5573	2.5154	0.0064
			95%	2.8650	2.8703	2.8492	0.0094
	D15	4.3951	90%	2.4939	2.4998	2.4406	0.0067
			95%	2.8135	2.8160	2.7865	0.0081
	D16	4.4047	90%	2.4947	2.5006	2.4883	0.0060
			95%	2.8141	2.8168	2.8012	0.0077
(-0.7, 0.7)	D13	9.0265	90%	2.7317	2.7870	2.7060	0.0071
			95%	3.0527	3.0885	3.0113	0.0088
	D14	1.6200	90%	2.1491	2.1491	2.1477	0.0069
			95%	2.4838	2.4839	2.4955	0.0094
	D15	0.9211	90%	1.9972	1.9972	1.9958	0.0066
			95%	2.3386	2.3385	2.3342	0.0089
	D16	3.1220	90%	2.3697	2.3707	2.3558	0.0069
			95%	2.6937	2.6939	2.7003	0.0088
(0, 2)	D17	9.1250	90%	2.7349	2.7915	2.7118	0.0064
			95%	3.0561	3.0927	3.0191	0.0097
	D18	5.4371	90%	2.5693	2.5832	2.5350	0.0068
			95%	2.8879	2.8949	2.8714	0.0089
	D19	4.4390	90%	2.4975	2.5036	2.2368	0.0065
			95%	2.8168	2.8197	2.5548	0.0100
	D20	4.4047	90%	2.4947	2.5006	2.4807	0.0066
			95%	2.8141	2.8168	2.8113	0.0091
(0.6, 1.4)	D17	7.7445	90%	2.6859	2.7250	2.6734	0.0061
			95%	3.0055	3.0294	2.9965	0.0082
	D18	0.4016	90%	1.8472	1.8473	1.8443	0.0077
			95%	2.1924	2.1924	2.1882	0.0077
	D19	0.2045	90%	1.7782	1.7782	1.7843	0.0071
			95%	2.1240	2.1239	2.1238	0.0094
	D20	1.6359	90%	2.1521	2.1521	2.1825	0.0068
			95%	2.4868	2.4867	2.5119	0.0078
(-2, 0)	D21	9.0585	90%	2.7327	2.7885	2.6984	0.0067
			95%	3.0539	3.0898	3.0313	0.0085
	D22	5.8540	90%	2.5947	2.6125	2.5742	0.0068
			95%	2.9132	2.9227	2.8829	0.0082
	D23	5.3951	90%	2.5667	2.5801	2.4705	0.0063
			95%	2.8851	2.8920	2.7948	0.0096
	D24	5.5182	90%	2.5744	2.5890	2.5779	0.0073
			95%	2.8930	2.9004	2.8764	0.0083
(-1.5, -0.5)	D21	6.3495	90%	2.6220	2.6450	2.5889	0.0069
			95%	2.9408	2.9534	2.9319	0.0092
	D22	0.9055	90%	1.9934	1.9933	1.9995	0.0073
			95%	2.3346	2.3347	2.3367	0.0095
	D23	2.0226	90%	2.2194	2.2193	2.2092	0.0066
			95%	2.5506	2.5505	2.5419	0.0087
	D24	2.5298	90%	2.2951	2.2952	2.3250	0.0063
			95%	2.6224	2.6225	2.6458	0.0090

Table 5.16: Critical values for 4th order polynomial regression

5.5.3 For bivariate linear regression

Also, it is of interest to compare the methods of constructing confidence bands for a multiple linear regression. Here, we consider the bivariate linear regression model, which is the simplest one in multiple case, with the predictor variables restricted in a rectangular region $\mathcal{X} \subset \mathcal{R}^2$. The methods concerned in this comparison are the approximate method of Sun and Loader (1994) and the simulation-based method of Liu, Jamshidian, Zhang and Donnelly (2005).

For the approximate method, we calculate critical values using the formula given in Proposition 5.2.2 that

$$\begin{aligned}\alpha &= \frac{\kappa_0}{\pi^{3/2}}\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)}\frac{c}{\sqrt{\nu}}(1+\frac{c^2}{\nu})^{-(\nu+1)/2} \\ &\quad +\frac{\zeta_0}{2\pi}(1+\frac{c^2}{\nu})^{-\nu/2}+\mathrm{P}\{|t_\nu|>c\}. \end{aligned} \tag{5.36}$$

Constants κ_0 and ζ_0 can be computed by

$$\kappa_0 = \int_{\mathcal{X}} \det{}^{1/2}(A^TA)d\mathbf{x}, \tag{5.37}$$

$$\zeta_0 = \int_{\partial\mathcal{X}} \det{}^{1/2}(\mathbf{A}_*^T\mathbf{A}_*), \tag{5.38}$$

where $A = (\mathrm{T}_1(\mathbf{x}), \mathrm{T}_2(\mathbf{x}))$, and $\mathbf{A}_* = \mathrm{T}_1(\mathbf{x})$ or $\mathrm{T}_2(\mathbf{x})$ with $\mathrm{T}_j(\mathbf{x})$ defined by $\mathrm{T}_j(\mathbf{x}) = \partial\mathrm{T}(\mathbf{x})/\partial x_j$ for $j = 1, 2$. Note that $\mathrm{T}(\mathbf{x}) = \mathrm{l}(\mathbf{x})/\|\mathrm{l}(\mathbf{x})\|$, where $\mathrm{l}(\mathbf{x}) = X(X^TX)^{-1}\mathbf{x}$, and $\mathbf{x} = (1, x_1, x_2)^T$. Thus, we have

$$\begin{aligned}\mathrm{T}_1(\mathbf{x}) &= \frac{\mathrm{l}'(\mathbf{x})\|\mathrm{l}(\mathbf{x})\| - \mathrm{l}(\mathbf{x})\|\mathrm{l}(\mathbf{x})\|'}{\|\mathrm{l}(\mathbf{x})\|^2} \\ &= \frac{\mathrm{l}'(\mathbf{x})(\|\mathrm{l}(\mathbf{x})\|^2)^{1/2} - \mathrm{l}(\mathbf{x})[(\|\mathrm{l}(\mathbf{x})\|^2)^{1/2}]'}{\|\mathrm{l}(\mathbf{x})\|^2} \\ &= \frac{\mathrm{l}'(\mathbf{x})(\|\mathrm{l}(\mathbf{x})\|^2)^{1/2} - (1/2)\mathrm{l}(\mathbf{x})(\|\mathrm{l}(\mathbf{x})\|^2)^{-1/2}(\|\mathrm{l}(\mathbf{x})\|^2)'}{\|\mathrm{l}(\mathbf{x})\|^2}, \end{aligned} \tag{5.39}$$

where

$$\begin{aligned}\mathrm{l}'(\mathbf{x}) &= \partial\mathrm{l}(\mathbf{x})/\partial x_1 = X(X^TX)^{-1}(0,1,0)^T, \\ \|\mathrm{l}(\mathbf{x})\|^2 &= \mathrm{l}^T(\mathbf{x})\mathrm{l}(\mathbf{x}) = \mathbf{x}^T(X^TX)^{-1}\mathbf{x}, \\ (\|\mathrm{l}(\mathbf{x})\|^2)' &= 2\cdot\mathrm{l}^T(\mathbf{x})\mathrm{l}'(\mathbf{x}) = 2\cdot\mathrm{l}^T(\mathbf{x})\cdot X(X^TX)^{-1}(0,1,0)^T. \end{aligned}$$

Similarly, we can obtain $T_2(x)$ by replacing $(0,1,0)^T$ by $(0,0,1)^T$ and $T_1(x)$ by $T_2(x)$ in (5.39). Then both A and A_* are ready. Numerical integrations may be used to compute constants κ_0 and ζ_0.

For the simulation-based method, as described in Section 5.4, the main task is to solve the optimization problem to find T. In practice, we compute critical values using ***SimReg*** software from Jamshidin, Liu, Zhang and Jamshidian (2004) on MATLAB 7 platform.

Now, we turn to choose the levels of the common factors for our comparison. Apparently, both methods depend on the design matrix, restricted intervals of the two predictor variables x_1 and x_2, and the confidence level. Furthermore, recall (5.28) that

$$T = Q\frac{\|\mathbf{N}\|}{(\hat{\sigma}/\sigma)},$$

where

$$Q = \sup_{x_1\in(a_1,b_1),x_2\in(a_2,b_2)} \frac{|(P\mathbf{x})^T\mathbf{N}|}{\|P\mathbf{x}\|\|\mathbf{N}\|}.$$

Since $\mathbf{N}$ and $(\hat{\sigma}/\sigma)$ are generated numbers, the simulation-based method, in fact, depends on the 3-dimensional vector $P\mathbf{x}$ together with the two restricted intervals (a_1,b_1) and (a_2,b_2). Let $P = (\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2)$ and define the set

$$\begin{aligned}\mathcal{L} &= \{P\mathbf{x} : \mathbf{x} \in \mathcal{X}\} \\ &= \{\mathbf{p}_0 + x_1\mathbf{p}_1 + x_2\mathbf{p}_2 : x_1 \in [a_1,b_1], x_2 \in [a_2,b_2]\}.\end{aligned}$$

Then, we have

$$Q = \sup_{\mathbf{v}\in\mathcal{L}} \frac{\mathbf{v}^T\mathbf{N}}{\|\mathbf{v}\|\|\mathbf{N}\|}, \tag{5.40}$$

where $\mathbf{v}^T\mathbf{N}/\|\mathbf{v}\|\|\mathbf{N}\|$ is simply the cosine of the angle between $\mathbf{v}$ and $\mathbf{N}$. This determines that, in order to obtain Q, it is sufficient to find the smallest angle between either $\mathbf{N}$ and $\mathbf{v}$ or $-\mathbf{N}$ and $\mathbf{v}$, as $\mathbf{v}$ ranges in $\mathcal{L}$. Note that $\mathcal{L}$ is the

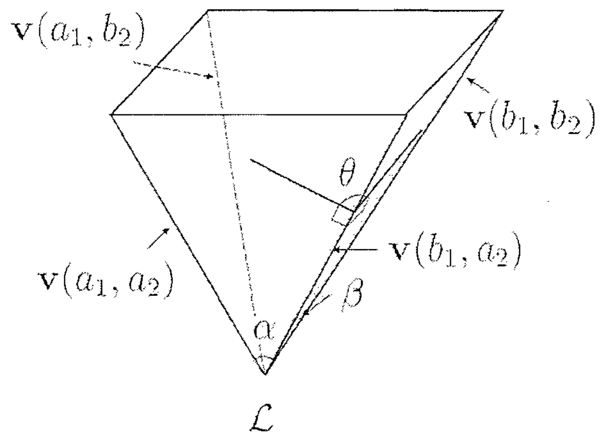

Figure 5.2: The cone determined by three angles

cone spanned by the following four vectors

$$\begin{aligned}
\mathbf{v}(a_1, a_2) &= \mathbf{p}_0 + a_1\mathbf{p}_1 + a_2\mathbf{p}_2, \\
\mathbf{v}(b_1, a_2) &= \mathbf{p}_0 + b_1\mathbf{p}_1 + a_2\mathbf{p}_2, \\
\mathbf{v}(a_1, b_2) &= \mathbf{p}_0 + a_1\mathbf{p}_1 + b_2\mathbf{p}_2, \\
\mathbf{v}(b_1, b_2) &= \mathbf{p}_0 + b_1\mathbf{p}_1 + b_2\mathbf{p}_2,
\end{aligned}$$

and it is depicted in Figure 5.2.

Define three angles that α is the angle between $\mathbf{v}(a_1, a_2)$ and $\mathbf{v}(b_1, a_2)$, β is the angle between $\mathbf{v}(b_1, a_2)$ and $\mathbf{v}(b_1, b_2)$, θ is the angle between the two planes S_1 and S_2, where S_1 is spanned by $\mathbf{v}(a_1, a_2)$ and $\mathbf{v}(b_1, a_2)$, S_2 is spanned by $\mathbf{v}(b_1, a_2)$ and $\mathbf{v}(b_1, b_2)$. Then the cone $\mathcal{L}$ is determined by these three angles. Therefore, we may conclude that the simulation-based method essentially depends on the three angles α, β, θ, the degree of freedom, and the confidence level. This is the nature of the simulation-based method. Also, the approximate method depends on the three angles via the design matrix and the two restricted intervals of the predictor variables. In this comparison, we fix a design matrix and then appropriately choose the restricted intervals such that their combinations determine manifold levels of the three angles.

First, we come to choose the design matrix. The acetylene data of Snee

(1977) is adopted here, since it was used to construct confidence bands in many papers. Another reason to select this dataset is that the critical value computed by my MATLAB programme of the approximate method is the same as the one given in Sun and Loader (1994) for this data when the confidence level is 95% and $\mathcal{X} = [1100, 1300] \times [5.3, 23]$. So we can reasonably think that my programme is reliable.

Next, we turn to choose the restricted intervals of the two predictor variables. First consider the three angles α, β and θ. We are willing to obtain various structures of the cone determined by these three angles. Set three levels for each angle from small to large within its range $[0, \pi]$. So we have $3 \times 3 \times 3 = 27$ structures for the cone. However, realize that for a fixed θ when α and β switch their values, the cone newly obtained has a similar structure with the original cone. Also, note that for fixed α and β, the cone with large θ is similarly structured with the one having small θ. Thus, for the sake of similarity, the cone finally has 12 structures of interest. They are SSM, SMM, SLM, MMM, MLM, LLM, SSL, SML, SLL, MML, MLL, LLL, where characters S, M, L stand for small value (around $\pi/12$), medium value (around $\pi/2$) and large value (around $8\pi/9$) respectively for the three angles, and the 3-character string denotes the level of each angle in order, e.g., SLM describes the situation that α is small, β is large, and θ is medium. Now, it is ready to choose the restricted intervals for predictor variables. We use the design matrix and any two initial guesses of restricted intervals to calculate α, β and θ, then adjust the restricted intervals to ensure the three angles each is near the level of our interest. One may follow the procedure below to find the restricted intervals:

Step 1 Adjust b_1 and a_2 to ensure θ is ok.

Step 2 Fix b_1 and a_2, adjust a_1 to ensure α is ok.

Step 3 Fix b_1 and a_2, adjust b_2 to ensure β is ok.

In such a way, we obtain 12 pairs of the restricted intervals for the predictor variables x_1 and x_2 according to 12 structures of the cone.

d.f. = 5

Restricted intervals	Three angles alpha	beta	theta	Confidence level	Critical values cv_app	cv_simu	s.e.
(10, 950)	0.2183	0.2315	1.6041	90%	2.1691	2.2749	0.0081
(13, 15)	S	S	M	95%	2.7441	2.8654	0.0122
(1070, 1250)	0.3071	1.5149	1.5630	90%	3.2006	3.1088	0.0098
(10, 50)	S	M	M	95%	3.8945	3.7947	0.0128
(920, 1620)	0.4079	2.7006	1.6081	90%	3.3834	3.1657	0.0089
(12, 130)	S	L	M	95%	4.0981	3.9022	0.0152
(100, 1220)	1.6414	1.6184	1.5686	90%	3.0510	2.9914	0.0082
(0, 12)	M	M	M	95%	3.7272	3.6839	0.0147
(650, 1600)	1.6860	2.8126	1.6059	90%	3.3808	3.1713	0.0088
(12, 40)	M	L	M	95%	4.0947	3.8897	0.0158
(10, 1600)	2.8843	2.8546	1.6059	90%	3.1580	2.9299	0.0090
(12, 15)	L	L	M	95%	3.8405	3.5917	0.0149
(1450, 1650)	0.2418	0.1853	2.8146	90%	2.3135	2.3478	0.0077
(-10, 5)	S	S	L	95%	2.9053	2.9487	0.0123
(950, 1800)	0.2485	1.6552	2.8438	90%	3.6521	3.2924	0.0104
(-10, 250)	S	M	L	95%	4.4000	4.0240	0.0137
(900, 1750)	0.2677	2.7231	2.9895	90%	3.2545	3.0841	0.0094
(-200, 10)	S	L	L	95%	3.9523	3.7844	0.0134
(950, 1800)	1.6552	1.6400	2.8438	90%	3.6341	3.2920	0.0105
(-10, 40)	M	M	L	95%	4.3798	4.0229	0.0137
(800, 2000)	1.6351	2.8342	2.9377	90%	3.2630	3.0733	0.0105
(-22, 10)	M	L	L	95%	3.9613	3.7748	0.0131
(400, 2200)	2.7204	2.8541	2.7026	90%	3.0603	2.8362	0.0084
(0, 5)	L	L	L	95%	3.7288	3.5086	0.0163

Table 5.17: Critical values for bivariate linear regression

We still use 90% and 95% confidence levels. Moreover, we manually set the degree of freedom equal to 5 and 30 to have a general view. Critical values are computed based on the designs described above. All the results are contained in Tables 5.17 and 5.18. Note that the critical values of the simulation-based method are based on 100,000 simulations.

From the result, we can draw some conclusions. The simulation-based method obtains smaller critical values than the approximate method generally, except for the cases when α and β are both small. When $\nu = 5$, the difference between the critical values of the two methods is generally around 0.2; the critical values of the approximate method are at most 11% larger than those of the simulation-based method. When $\nu = 30$, the difference between the critical values of the two methods is not apparent, generally at the second decimal place.

d.f. = 30

Restricted intervals	Three angles alpha	beta	theta	Confidence level	Critical values cv_app	cv_simu	s.e.
(10, 950)	0.2183	0.2315	1.6041	90%	1.8105	1.8808	0.0052
(13, 15)	*S*	*S*	*M*	95%	2.1566	2.2243	0.0066
(1070, 1250)	0.3071	1.5149	1.5630	90%	2.4886	2.4657	0.0051
(10, 50)	*S*	*M*	*M*	95%	2.8211	2.8037	0.0066
(920, 1620)	0.4079	2.7006	1.6081	90%	2.5937	2.5257	0.0047
(12, 130)	*S*	*L*	*M*	95%	2.9223	2.8666	0.0066
(100, 1220)	1.6414	1.6184	1.5686	90%	2.3970	2.3846	0.0049
(0, 12)	*M*	*M*	*M*	95%	2.7318	2.7247	0.0064
(650, 1600)	1.6860	2.8126	1.6059	90%	2.5911	2.5158	0.0046
(12, 40)	*M*	*L*	*M*	95%	2.9196	2.8579	0.0066
(10, 1600)	2.8843	2.8546	1.6059	90%	2.4485	2.3550	0.0055
(12, 15)	*L*	*L*	*M*	95%	2.7764	2.6967	0.0071
(1450, 1650)	0.2418	0.1853	2.8146	90%	1.9131	1.9404	0.0050
(-10, 5)	*S*	*S*	*L*	95%	2.2583	2.2829	0.0066
(950, 1800)	0.2485	1.6552	2.8438	90%	2.7480	2.6028	0.0051
(-10, 250)	*S*	*M*	*L*	95%	3.0725	2.9533	0.0063
(900, 1750)	0.2677	2.7231	2.9895	90%	2.5146	2.4507	0.0048
(-200, 10)	*S*	*L*	*L*	95%	2.8443	2.7920	0.0068
(950, 1800)	1.6552	1.6400	2.8438	90%	2.7382	2.6016	0.0051
(-10, 40)	*M*	*M*	*L*	95%	3.0630	2.9528	0.0064
(800, 2000)	1.6351	2.8342	2.9377	90%	2.5183	2.4453	0.0047
(-22, 10)	*M*	*L*	*L*	95%	2.8474	2.7861	0.0069
(400, 2200)	2.7204	2.8541	2.7026	90%	2.3842	2.2862	0.0050
(0, 5)	*L*	*L*	*L*	95%	2.7111	2.6270	0.0069

Table 5.18: Critical values for bivariate linear regression

5.5.4 Conclusions

Overall, it may be concluded from our comparisons that the simulation-based method computes as good critical values as the exact method, better than either the conservative method or the approximate method. When we increase the number of simulations, the simulation-based method may get even accurate critical values. Meanwhile, it can be found that Naiman's method is good enough. That is because Naiman's critical values are for conservative confidence bands but they are actually not much conservative. Comparatively speaking, the approximate method is bad, but not seriously.

5.6 Numerical examples

5.6.1 Example for simple linear regression

In an 1857 article, a Scottish physicist named James D. Forbes discussed a series of experiments that he had done concerning the relationship between atmospheric pressure and the boiling point of water. He believed that altitude could be determined by atmospheric pressure, measured with a barometer which was a fragile instrument in the middle of the nineteenth century, with lower pressures corresponding to higher altitudes. Forbes wondered whether a simpler measurement of the boiling point of water could substitute for a direct reading of barometric pressure to determine the altitude. He collected data in the Alps and in Scotland and measured pressure in inches of mercury with barometer and boiling point in degrees Fahrenheit using a thermometer at each location. Boiling point measurements were adjusted for the difference between the ambient air temperature when he took the measurements and a standard temperature. The data for 17 locations are reproduced in Table 5.19, which is taken from Weisberg (2005, page 22).

A simple linear regression model is used to fit the data. Atmospheric pressure is viewed as the response and the boiling point of water is regarded as the only predictor variable in the model. Therefore, we have the fitted

Case Number	Temperature	Pressure
1	194.5	20.79
2	194.3	20.79
3	197.9	22.40
4	198.4	22.67
5	199.4	23.15
6	199.9	23.35
7	200.9	23.89
8	201.1	23.99
9	201.4	24.02
10	201.3	24.01
11	203.6	25.14
12	204.6	26.57
13	209.5	28.49
14	208.6	27.76
15	210.7	29.04
16	211.9	29.88
17	212.2	30.06

Table 5.19: Forbes' 1857 data on boiling point and barometric pressure for 17 locations in the Alps and Scotland

regression model given by

$$\hat{y} = -81.0637 + 0.5229x. \tag{5.41}$$

Simultaneous confidence bands can be constructed then over a restricted interval, say, [194.3, 212.2] which takes the smallest and largest observations as the lower and upper bounds. The exact method provides critical values 2.2822, 2.6693 and 3.5122 for 90%, 95% and 99% confidence levels respectively; the approximate method suggests 2.3171, 2.6946, 3.5270, and the simulation-based method gives 2.2837, 2.6715, 3.4968 correspondingly. Note that the simulation-based method is on a basis of 100,000 replicates, and will be so for the other examples in this chapter. The confidence bands are plotted in Figures 5.3-5.5.

5.6.2 Example for polynomial regression

Table 5.20 presents data concerning the strength of kraft paper and the percentage of hardwood in the batch of pulp from which the paper was produced.

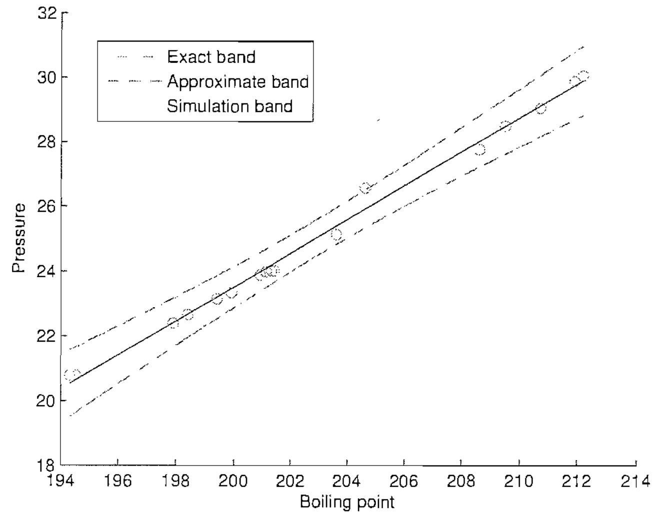

Figure 5.3: Confidence bands for 90% confidence level

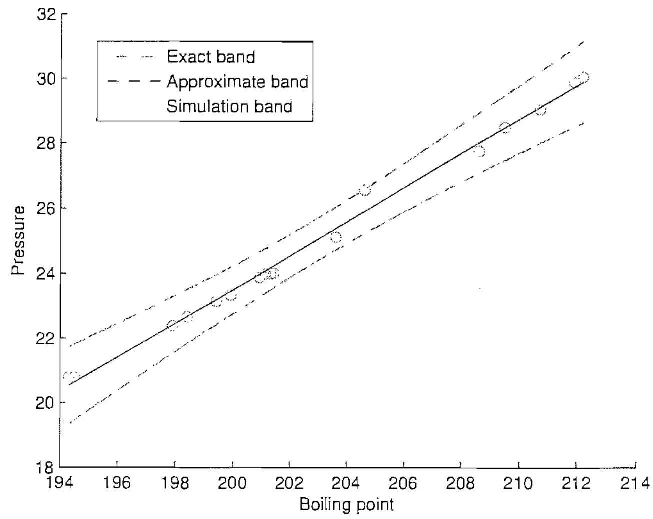

Figure 5.4: Confidence bands for 95% confidence level

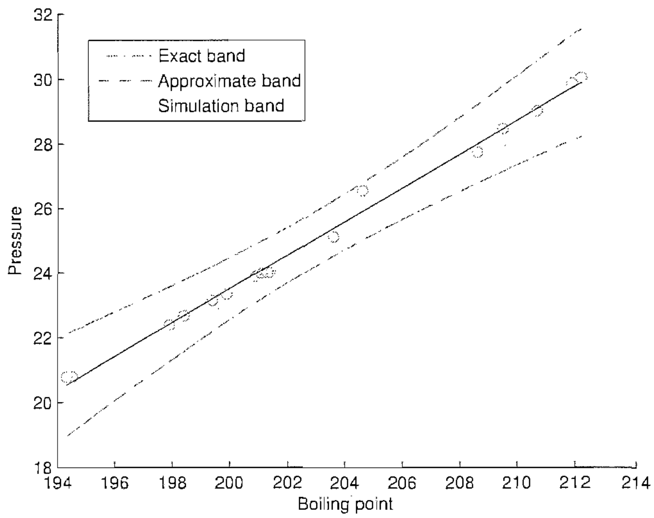

Figure 5.5: Confidence bands for 99% confidence level

These data is taken from Montgomery, Peck and Vining (2006, page 205).

A scatter plot of the data displays that a quadratic regression model may adequately describe the relationship between tensile strength and hardwood concentration. According to these data, the fitted model is given by

$$\hat{y} = -6.6742 + 11.7640x - 0.6345x^2. \tag{5.42}$$

Note that $x\%$ here is a percentage so that x should be bounded by the interval [0, 100]. Then we construct simultaneous confidence bands over this restricted interval using Naiman's methods, the approximate method, and the simulation-based method. Consequently, these three methods give critical values 2.5661, 2.6476, 2.5483 for 90% confidence level, and 2.9482, 3.0095, 2.9396 for 95% level, respectively. We plot the confidence bands in Figures 5.6 and 5.7. Note that the bands plotted in the figures are parts of the whole bands over the restricted interval. Doing this is in order to make the observed points more clear.

Hardwood concentration (%)	Tensile strength (psi)
1	6.3
1.5	11.1
2	20.0
3	24.0
4	26.1
4.5	30.0
5	33.8
5.5	34.0
6	38.1
6.5	39.9
7	42.0
8	46.1
9	53.1
10	52.0
11	52.5
12	48.0
13	42.8
14	27.8
15	21.9

Table 5.20: Hardwood concentration in pulp and tensile strength of kraft paper

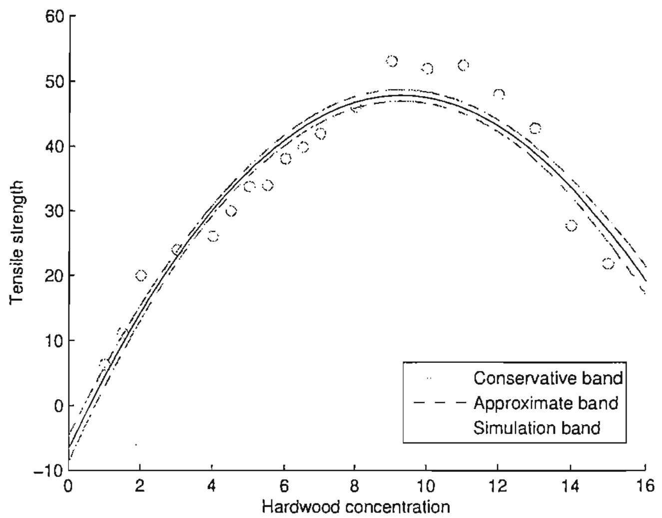

Figure 5.6: Confidence bands for 90% confidence level

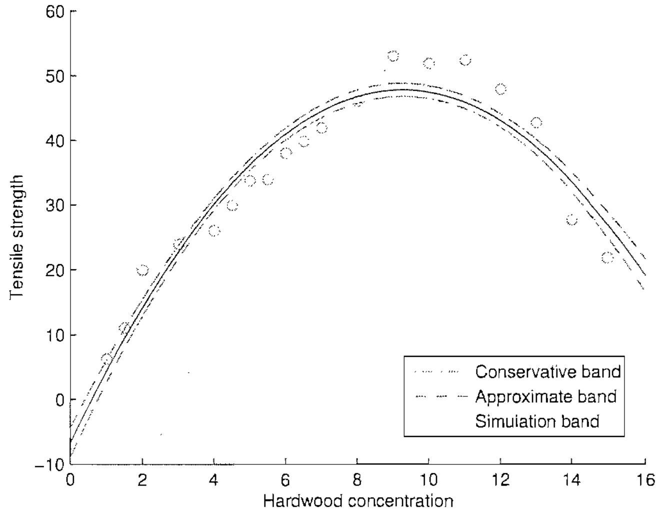

Figure 5.7: Confidence bands for 95% confidence level

5.6.3 Example for bivariate linear regression

A soft drink bottler is analyzing the vending machine service routes in his distribution system. He is interested in predicting the amount of time required by the route driver to service the vending machines in an outlet. This service activity includes stocking the machine with beverage products and minor maintenance or housekeeping. It is suggested by the industrial engineer for this study that the two most important factors affecting the delivery time (y) are the number of cases of product stocked (x_1) and the distance walked by the route driver (x_2). 25 observations on the delivery time has been collected by the engineer, and they are shown in Table 5.21. These data is also taken from Montgomery, Peck and Vining (2006, page 70).

We fit the data using a bivariate linear regression model. Therefore, the fitted model is given by

$$\hat{y} = 2.3412 + 1.6159x_1 + 0.0144x_2. \tag{5.43}$$

We assume the maximum capacity of product stocked is 30 cases and the distance is preferred within 2000 ft. Then x_1 and x_2 should be bounded by

Observation number	Delivery Time (min)	Number of cases	Distance (ft)
1	16.68	7	560
2	11.50	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8.00	2	110
8	17.83	7	210
9	79.24	30	1460
10	21.50	5	605
11	40.33	16	688
12	21.00	10	215
13	13.50	4	255
14	19.75	6	462
15	24.00	9	448
16	29.00	10	776
17	15.35	6	200
18	19.00	7	132
19	9.50	3	36
20	35.10	17	770
21	17.90	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

Table 5.21: Delivery time data for bivariate example

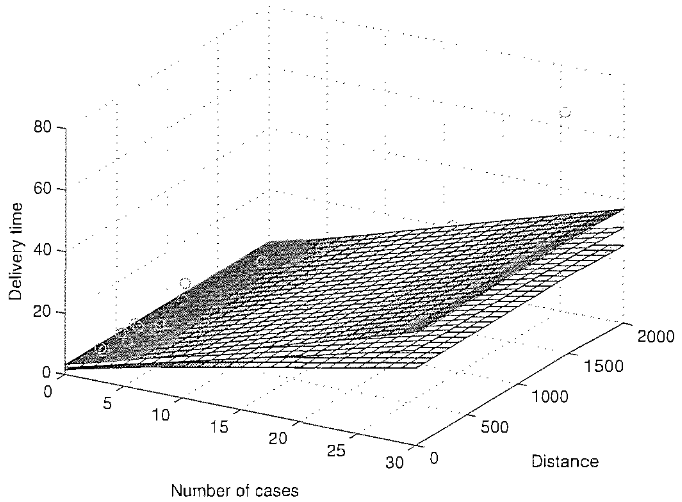

Figure 5.8: The approximate band for 90% confidence level

the intervals [0, 30] and [0, 2000] respectively. In such a case, simultaneous confidence bands can be constructed accordingly based on the approximate method and the simulation-based method. As results, these two methods suggest critical values 2.7234, 2.6409 for 90% confidence level, and 3.0707, 2.9787 for 95% confidence level, respectively. To be clear, we plot single band in each picture. So the four confidence bands are shown in Figures 5.8-5.11 respectively.

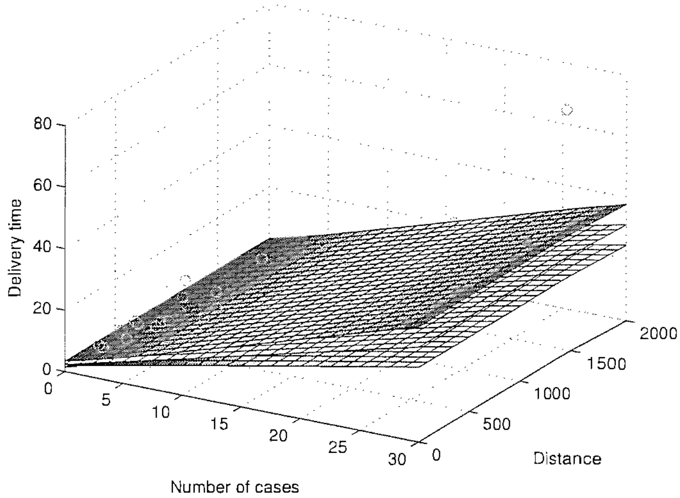

Figure 5.9: The approximate band for 95% confidence level

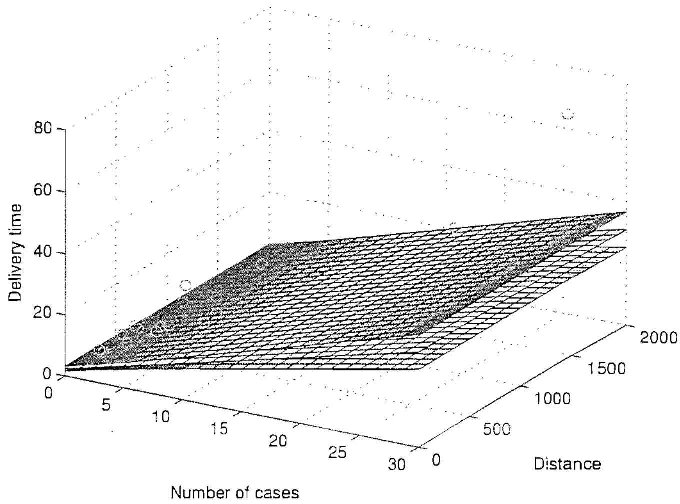

Figure 5.10: The simulation-based band for 90% confidence level

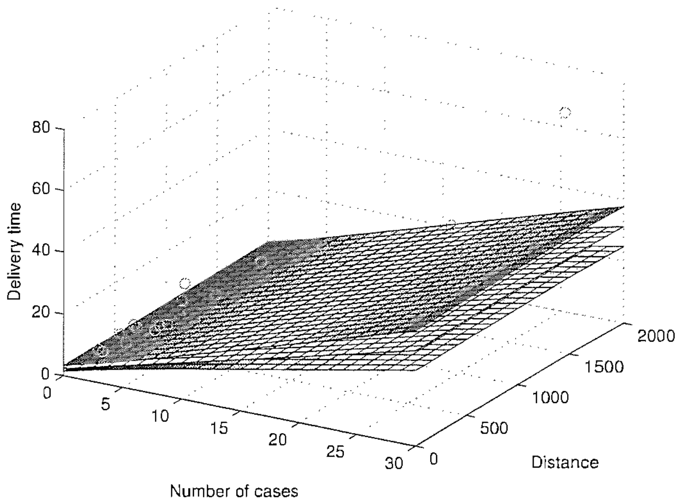

Figure 5.11: The simulation-based band for 95% confidence level

Chapter 6

Simultaneous confidence bands for a logistic regression model

The analysis of dichotomous response data has been popular due to the increasing use of the logistic regression model which enjoys a wide variety of applications nowadays, such as medical treatment, clinical trials, epidemiological test and risk management. Construction of simultaneous confidence bands for a logistic regression model is therefore of interest. However, the existing literatures on this are very limited.

Since the asymptotic distributional approximation of the parameter estimators of interest is the base of construction of confidence regions for a generalized linear model, we first briefly review relevant literatures on, for example, the construction of asymptotic intervals for the binomial parameter. By recalling related works, a general profile on the quality of the asymptotic approximation based on several methods would be obtained. The most frequently mentioned interval in many statistical textbooks is the standard or Wald confidence interval. This interval was shown to perform poorly unless the sample size is quite large in, e.g., Ghosh (1979), Blyth and Still (1983). Clopper and Pearson (1934) proposed "exact" confidence interval based on inverting equal-tailed binomial tests. The "exact" interval is usually necessarily conservative. Therefore, it is inappropriate to treat it as optimal for statistical practice. Agresti and Coull (1998) discussed that the score confidence interval first presented by Wilson (1927) tends to perform much better than the Wald or "exact" intervals in terms of having coverage probabilities close to the nominal confidence level. Zheng and Loh (1995) and Zheng (1998) considered bootstrapping binomial confidence intervals via bootstrap calibration. Brown, Cai and DasGupta (2000, 2001, 2002) provided a survey of these intervals as well as the Bayes credible intervals, and gave comparisons. Chen (1990) demonstrated the accuracy of such approximate intervals for a binomial parameter.

For the construction of simultaneous confidence bands, an alternative of the methods based on the asymptotic distributional property is the bootstrap percentile method, which was proposed by Yeh (1996) to construct confidence bands for unknown curves based on the bootstrap and the concept of "curve depth". However, it is not considered further in this thesis.

In this chapter, we first introduce some key methods of constructing two-sided simultaneous confidence bands for the probability of the dichotomous response in a logistic regression model with or without constrained predictor variables. Two examples are given for one-dimensional and two-dimensional cases respectively. Then simulation studies are given for the comparison of the methods. Meanwhile, the simulation results can be used to gauge the accuracy of the asymptotic distributional approximation. That is to check whether the simulated coverage probability of certain band is close to the nominal confidence level, and how far between them.

6.1 Confidence bands for a logistic regression without constraint on predictor variables

6.1.1 For a simple logistic regression

Brand, Pinnock and Jackson (1973) described a method of constructing large sample confidence bands for the logistic response curve for the case of $p = 1$, where p is the number of the predictor variables in the regression model.

A data set in this case consists of pairs $(x_i, y_i), i = 1, 2, 3, \ldots, N$, where x_i is, say, a measure of dose received by the ith test subject and y_i is set to 1 or 0 respectively corresponding to that the response does or does not occur in the ith subject. N is required sufficiently large for large sample normality to be a reasonable approximation.

The probability of the response corresponding to dose x, $0 \leq \pi(x) \leq 1$, is defined in terms of parameters β_0 and β_1 as

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}. \tag{6.1}$$

Suppose $\hat{\beta}_0$ and $\hat{\beta}_1$ are the maximum likelihood estimators of parameter β_0 and β_1. Thus the components of the information matrix I_{11}, I_{12}, I_{22} can be expressed by

$$\begin{aligned}
I_{11} &= \sum_{i=1}^{N} \{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)/[1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2\}, \\
I_{12} &= \sum_{i=1}^{N} \{x_i \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)/[1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2\}, \\
I_{22} &= \sum_{i=1}^{N} \{x_i^2 \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)/[1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2\}.
\end{aligned}$$

Recall the large sample asymptotic normality of $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)^T$ given by

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N_2(\mathbf{0}, \Sigma), \tag{6.2}$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, N_2 denotes the bivariate normal distribution, and Σ is the asymptotic covariance matrix. Note that the asymptotic covariance matrix for $\hat{\boldsymbol{\beta}}$, Σ/N, can be estimated by the inverse of the information matrix J^{-1}, i.e., $\Sigma \approx NJ^{-1}$. Therefore, we have

$$\begin{aligned}
& N(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \Sigma^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
\approx\ & (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T J (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
=\ & I_{11}(\hat{\beta}_0 - \beta_0)^2 + 2I_{12}(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) + I_{22}(\hat{\beta}_1 - \beta_1)^2 \xrightarrow{\mathcal{D}} \chi_2^2, (6.3)
\end{aligned}$$

where χ_2^2 is the Chi-square distribution with two degrees of freedom. A large sample $1-\alpha$ confidence set for (β_0, β_1) is therefore given by

$$I_{11}(\hat{\beta}_0 - \beta_0)^2 + 2I_{12}(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1) + I_{22}(\hat{\beta}_1 - \beta_1)^2 \leq \chi^2_{2,\alpha}, \quad (6.4)$$

where $\chi^2_{2,\alpha}$ denotes the upper α percentage point of the χ_2^2 distribution. Note that the confidence set given by (6.4) is an ellipse centered at $(\hat{\beta}_0, \hat{\beta}_1)$. A $100(1-\alpha)\%$ level confidence band for $\pi(x)$ can be obtained by finding the maximal and minimal values of $\pi(x)$ over the confidence ellipse of (β_0, β_1) for each x.

Transform (6.1) to the form of interest

$$\lambda(x) = \beta_0 + \beta_1 x = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) \quad (6.5)$$

which is a monotone function of $\pi(x)$. We can equivalently find the extremal values of $\lambda(x)$ over the confidence ellipse (6.4) of (β_0, β_1) for each x.

For a fixed x, $\{\beta_0 + \beta_1 x = q : -\infty < q < \infty\}$ stands for a family of parallel straight lines in the (β_0, β_1)-plane with x as the slope. Extremal values of $\lambda(x) = \beta_0 + \beta_1 x$ over the confidence set (6.4) of (β_0, β_1) are attained when the lines in the family are tangent to the boundary of the confidence ellipse (6.4). Each tangent line corresponds to one (β_0, β_1) which can be viewed as a solution to the pair of equation (6.5) and the equality obtained by changing the sign from "$\leq$" to "$=$" in (6.4). Expressing $\beta_0 = \lambda(x) - \beta_1 x$ from (6.5) and substituting in the obtained equality gives a quadratic equation in terms of β_1, which has the form given by

$$a\beta_1^2 + b\beta_1 + c = 0$$

with

$$\begin{aligned} a &= a_1, \\ b &= b_1 + b_2\lambda(x), \\ c &= c_1 + c_2\lambda(x) + c_3\lambda(x)^2, \end{aligned}$$

where

$$\begin{aligned}
a_1 &= I_{11}x^2 - 2I_{12}x + I_{22}, \\
b_1 &= 2I_{11}x\hat{\beta}_0 + 2I_{12}(\hat{\beta}_1 x - \hat{\beta}_0) - 2I_{22}\hat{\beta}_1, \\
b_2 &= -2I_{11}x + 2I_{12}, \\
c_1 &= I_{11}\hat{\beta}_0^2 + 2I_{12}\hat{\beta}_0\hat{\beta}_1 + I_{22}\hat{\beta}_1^2 - \chi^2_{2,\alpha}, \\
c_2 &= -2I_{11}\hat{\beta}_0 - 2I_{12}\hat{\beta}_1, \\
c_3 &= I_{11}.
\end{aligned}$$

Notice that there is only one solution of β_1 for a straight line tangent to the confidence ellipse. We have

$$b^2 - 4ac = 0,$$

which gives a quadratic equation in terms of $\lambda(x)$. And the resulting roots provide the extremal values of $\lambda(x)$ over the confidence set (6.4) of (β_0, β_1). We denote the maximum and minimum values of $\lambda(x)$ by $\lambda_H(x)$ and $\lambda_L(x)$ respectively, which are therefore given by

$$\begin{aligned}
\lambda_H(x), \lambda_L(x) &= \frac{-(2b_1b_2 - 4a_1c_2)}{2(b_2^2 - 4a_1c_3)} \\
&\pm \frac{[(2b_1b_2 - 4a_1c_3)^2 - 4(b_2^2 - 4a_1c_3)(b_1^2 - 4a_1c_1)]^{\frac{1}{2}}}{2(b_2^2 - 4a_1c_3)}. \qquad (6.6)
\end{aligned}$$

Also, $\lambda_H(x)$ and $\lambda_L(x)$ can be written in matrix form as

$$\begin{aligned}
\lambda_H(\mathbf{x}) &= \mathbf{x}^T\hat{\boldsymbol{\beta}} + (\chi^2_{2,\alpha}\mathbf{x}^T J^{-1}\mathbf{x})^{\frac{1}{2}}, \\
\lambda_L(\mathbf{x}) &= \mathbf{x}^T\hat{\boldsymbol{\beta}} - (\chi^2_{2,\alpha}\mathbf{x}^T J^{-1}\mathbf{x})^{\frac{1}{2}},
\end{aligned}$$

where $\mathbf{x} = (1, x)^T$ and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)^T$.

Then the confidence band for the probability of the dose-response with $100(1-\alpha)\%$ confidence level is given by

$$\left(\frac{\exp[\lambda_L(x)]}{1 + \exp[\lambda_L(x)]}, \frac{\exp[\lambda_H(x)]}{1 + \exp[\lambda_H(x)]}\right) \qquad \text{for all } x. \qquad (6.7)$$

6.1.2 For a multiple logistic regression

Hauck (1983) considered the construction of simultaneous confidence bands for the logistic response function with any number of predictor variables. He presented a computationally easier and more general method than that in Brand, Pinnock and Jackson (1973).

Let y be a dichotomous response with possible values 1 or 0. The probability of $y = 1$ denoted by $\pi(\mathbf{x})$, which is in terms of the predictor vector $\mathbf{x} = (x_1, \ldots, x_p)^T$, is given by

$$\pi(\mathbf{x}) = \frac{\exp(\mathbf{x}^T\boldsymbol{\beta})}{1 + \exp(\mathbf{x}^T\boldsymbol{\beta})}, \tag{6.8}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is the vector of the regression coefficients. Alternatively, (6.8) can be transformed as

$$\lambda(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta} = \ln\big(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\big).$$

Now, we assume observations $\mathbf{x}$ and y are of sample size N, which is large enough for the asymptotic normality of the maximum likelihood estimator vector $\hat{\boldsymbol{\beta}}$ to be a good approximation, i.e.,

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N_{p+1}(\mathbf{0}, \Sigma), \tag{6.9}$$

where Σ/N can be estimated by J^{-1} with J being the information matrix of $\hat{\boldsymbol{\beta}}$, which has the elements given by

$$\begin{aligned} J_{jk} &= \sum_i^N \hat{\pi}(\mathbf{x}_i)[1 - \hat{\pi}(\mathbf{x}_i)]x_{ij}x_{ik} \\ &= \sum_i^N \{\exp(\mathbf{x}^T\hat{\boldsymbol{\beta}})/[1 + \exp(\mathbf{x}^T\hat{\boldsymbol{\beta}})]^2\}x_{ij}x_{ik}, \end{aligned} \tag{6.10}$$

$i = 1, \ldots, N; j, k = 1, \ldots, p$. We subsequently have

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T J(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \chi^2_{p+1},$$

where J can be obtained from most statistical package directly. Let $\chi^2_{p+1,\alpha}$ denotes the upper α percentage point of the χ^2_{p+1} distribution, then

$$\mathrm{P}\{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T J(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \le \chi^2_{p+1,\alpha}\} \approx 1 - \alpha. \tag{6.11}$$

An approximate $1-\alpha$ confidence set for $\boldsymbol{\beta}$ is therefore given by

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T J(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \leq \chi^2_{p+1,\alpha}, \tag{6.12}$$

which is an ellipsoidal region centered at $\hat{\boldsymbol{\beta}}$.

Recall the Cauchy-Schwartz inequality which is of the form given by

$$|\mathbf{a}^T\mathbf{b}|^2 \leq \|\mathbf{a}\|^2 \cdot \|\mathbf{b}\|^2$$

for all vectors $\mathbf{a}$ and $\mathbf{b}$ in the inner product space. Applying the Cauchy-Schwartz inequality, we have for all $\mathbf{x}$,

$$\begin{aligned}
[\mathbf{x}^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]^2 &= |\mathbf{x}^T J^{-\frac{1}{2}} J^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})|^2 \\
&= \|[(J^{-\frac{1}{2}})^T\mathbf{x}]^T[J^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]\|^2 \\
&\leq \|(J^{-\frac{1}{2}})^T\mathbf{x}\|^2 \cdot \|J^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\|^2 \\
&= [(J^{-\frac{1}{2}})^T\mathbf{x}]^T[(J^{-\frac{1}{2}})^T\mathbf{x}] \cdot [J^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]^T[J^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})] \\
&= (\mathbf{x}^T J^{-1}\mathbf{x})[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T J(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]. \qquad (6.13)
\end{aligned}$$

Substitute (6.13) into (6.11) to obtain

$$\begin{aligned}
1-\alpha &\approx \mathrm{P}\{[\mathbf{x}^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]^2/[\mathbf{x}^T J^{-1}\mathbf{x}] \leq \chi^2_{p+1,\alpha}, \text{for all } \mathbf{x}\} \\
&= \mathrm{P}\{|\mathbf{x}^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})| \leq (\chi^2_{p+1,\alpha}\mathbf{x}^T J^{-1}\mathbf{x})^{\frac{1}{2}}, \text{for all } \mathbf{x}\} \\
&= \mathrm{P}\{\mathbf{x}^T\boldsymbol{\beta} \leq \mathbf{x}^T\hat{\boldsymbol{\beta}} \pm (\chi^2_{p+1,\alpha}\mathbf{x}^T J^{-1}\mathbf{x})^{\frac{1}{2}}, \text{for all } \mathbf{x}\}. \qquad (6.14)
\end{aligned}$$

Therefore a $1-\alpha$ level confidence band for $\mathbf{x}^T\boldsymbol{\beta}$ is given by

$$\begin{aligned}
&\big(\lambda_L(\mathbf{x}), \lambda_U(\mathbf{x})\big) \\
&= \big(\mathbf{x}^T\hat{\boldsymbol{\beta}} - (\chi^2_{p+1,\alpha}\mathbf{x}^T J^{-1}\mathbf{x})^{\frac{1}{2}}, \mathbf{x}^T\hat{\boldsymbol{\beta}} + (\chi^2_{p+1,\alpha}\mathbf{x}^T J^{-1}\mathbf{x})^{\frac{1}{2}}\big) \qquad (6.15)
\end{aligned}$$

for all $\mathbf{x}$. By making use of the logistic relationship, the corresponding confidence band for $\pi(\mathbf{x})$ is given by

$$\Big(\frac{\exp[\lambda_L(\mathbf{x})]}{1+\exp[\lambda_L(\mathbf{x})]}, \frac{\exp[\lambda_U(\mathbf{x})]}{1+\exp[\lambda_U(\mathbf{x})]}\Big) \quad \text{for all } \mathbf{x}. \tag{6.16}$$

6.2 Confidence bands for a logistic regression with restricted predictor variables

In real problems, some constraints may be imposed on the predictor variables. In this case, a confidence band is restricted to some subset of possible x's. Therefore, the bands described in section 6.1 are unnecessarily wide and conservative. Naturally, it is of great interest to consider methods of constructing confidence bands for a logistic regression with restricted predictor variables. In this section, we focus on this problem. The restricted predictor space considered is the most popular rectangular region, which has the form

$$\mathcal{X} = \{\mathrm{x} = (x_1, \dots, x_p), a_i \le x_i \le b_i, i = 1, \dots, p\}, \tag{6.17}$$

where a_i, b_i's are given real constants.

6.2.1 For a simple logistic regression

Band based on the method of Wynn and Bloomfield (1971)

For simple logistic regression, we have the asymptotic property that $\hat{\beta}$ is approximately normally distributed with mean β and estimated covariance matrix J^{-1} when the sample size N is sufficiently large. Then a confidence band can be constructed for $\mathrm{x}^T\beta$ of the form

$$\mathrm{x}^T\beta \in \mathrm{x}^T\hat{\beta} \pm c\sqrt{\mathrm{x}^T J^{-1}\mathrm{x}^T} \text{ all } x \in [a, b], \tag{6.18}$$

which can be written alternatively as

$$\sup_{x\in[a,b]} \frac{|\mathrm{x}^T(\hat{\beta} - \beta)|}{\sqrt{\mathrm{x}^T J^{-1}\mathrm{x}}} \le c, \tag{6.19}$$

where x is the only predictor variable which is restricted in the interval $[a, b]$, c is the critical value such that the band has the simultaneous coverage probability of $1 - \alpha$.

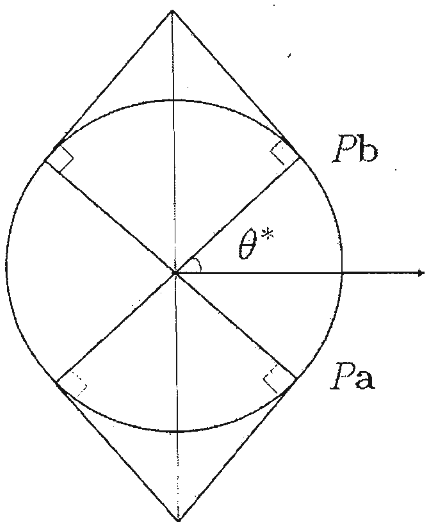

Figure 6.1: For Wynn and Bloomfield's method

Assume that there is a 2×2 non-singular matrix P such that $P^T P = J^{-1}$. Then (6.19) can be further written as

$$\sup_{x \in [a,b]} \frac{|(P\mathrm{x})^T (P^{-1})^T (\hat{\beta} - \beta)|}{\sqrt{(P\mathrm{x})^T (P\mathrm{x})}} = \sup_{x \in [a,b]} \frac{|(P\mathrm{x})^T \mathrm{N}|}{\|(P\mathrm{x})\|} \leq c, \qquad (6.20)$$

where $\mathrm{N} = (P^{-1})^T(\hat{\beta} - \beta)$ has a bivariate standard normal distribution, and so $\|\mathrm{N}\|^2$ has the χ_2^2 distribution.

Now, we turn to evaluate the probability of the event in (6.20) so as to find a computational formula for calculation of the critical value c. Define $\mathbf{a} = (1, a)^T, \mathbf{b} = (1, b)^T$, and θ^* is half the angle between $P\mathbf{a}$ and $P\mathbf{b}$. Figure 6.1 may be useful to easily calculate the last supreme in (6.20). $P\mathrm{x}$ is a vector moving within the circular cone bounded by $P\mathbf{a}$ and $P\mathbf{b}$, whereas N is a vector that can freely locate at any position in the plane. When N is within the circular cone or its opposite, the supreme is equal to $\|\mathrm{N}\|$; otherwise it equals the the projection of N on the nearest bound of the cones.

Assume θ is the angle between the positive horizontal axis and N. As the picture is symmetric, it is only needed to consider the part for $\theta \in [0, \pi/2]$.

We have

$$\sup_{x\in[a,b]} \frac{|(P\mathbf{x})^T\mathbf{N}|}{\|P\mathbf{x}\|} = \begin{cases} \|\mathbf{N}\| & \text{for } \theta \in [0,\theta^*], \\ \|\mathbf{N}\|\cos(\theta-\theta^*) & \text{for } \theta \in (\theta^*, \frac{\pi}{2}]. \end{cases}$$

The probability of the confidence band (6.18) is therefore equal to

$$\begin{aligned} & \mathrm{P}\{\sup_{x\in[a,b]} \frac{|(P\mathbf{x})^T\mathbf{N}|}{\|P\mathbf{x}\|} \le c\} \\ = \; & 4\big(\mathrm{P}\{\|\mathbf{N}\| \le c, 0 \le \theta \le \theta^*\} \\ & + \mathrm{P}\{\|\mathbf{N}\|\cos(\theta-\theta^*) \le c, \theta^* < \theta \le \frac{\pi}{2}\}\big) \\ = \; & 4\big(\frac{\theta^*}{2\pi}\chi_2^2(c^2) + \frac{1}{2\pi}\int_{\theta^*}^{\frac{\pi}{2}} \chi_2^2\big(\frac{c^2}{\cos^2(\theta-\theta^*)}\big)\,d\theta\big) \\ = \; & \frac{2\theta^*}{\pi}\chi_2^2(c^2) + \frac{2}{\pi}\int_0^{\frac{\pi}{2}-\theta^*} \chi_2^2\big(\frac{c^2}{\cos^2\theta}\big)\,d\theta, \end{aligned} \tag{6.21}$$

where θ^* can be calculated via the following formula

$$\theta^* = \frac{1}{2}\arccos\big(\frac{\mathbf{a}^T J^{-1}\mathbf{b}}{(\mathbf{a}^T J^{-1}\mathbf{a}\cdot\mathbf{b}^T J^{-1}\mathbf{b})^{1/2}}\big), \tag{6.22}$$

where J can be obtained directly from most statistical packages.

Consequently, given a confidence level, critical value c can be calculated from (6.21) and (6.22), which is used to construct confidence band (6.18) for $\mathbf{x}^T\boldsymbol{\beta}$. Hence, a confidence band can be obtained for the logistic response $\pi(x)$ by making use of the logistic relationship. This method is from Wynn and Bloomfield (1971), so we call the band of this method WB band hereafter.

Type 4 band of Sun, Loader and McCormick (2000)

Sun, Loader and McCormick (2000) considered confidence bands for generalized linear models. In their paper, it is stated that the approximation to the coverage probability of simultaneous confidence bands for the mean response function in linear models is still applicable without any change to the generalized linear models. However, in generalized linear models, the errors are often non-additive and non-normal. This may influence the accuracy of the approximation when the sample size is not large enough. Under this situation, they proposed to use an Edgeworth expansion for the distribution of $\hat{\beta}$ in connection with the idea of the Skorohod construction to convert an error term in the Edgeworth expansion to a bias term; then estimate and correct it to adjust the approximation formula such that the coverage probability of the corrected confidence band is much closer to the nominal confidence level. The correction proposed in the paper has two components: one is to apply the tube formula to some modified process; the other uses the method of bias correction in Sun and loader (1994).

For simplicity, only one-dimensional case was studied in Sun, Loader and McCormick (2000) but their method may be applied to cases of multiple dimension. They recommended their Type 4 confidence band which is given by

$$\mathbf{x}^T\hat{\beta} \pm (c - |\hat{r}_p|)(\mathbf{x}^T J^{-1}\mathbf{x})^{\frac{1}{2}} \quad \text{for all } x \in [a, b], \tag{6.23}$$

where c and $\hat{r}_p$ are a critical value and a corrected constant respectively. Their other types of bands are proven not to perform as well as Type 4 band for the logistic regression model when the sample size $N \geq 200$. Note that $(c - |\hat{r}_p|)$ in (6.23) as a whole can be obtained directly by using their software `parfit`, which can be downloaded from `www.locfit.info/`.

Note that the band in (6.23) is for $\mathbf{x}^T\beta$. The band for the logistic response $\pi(x)$ can be obtained from the band (6.23) in the usual way.

A numerical example

Consider the example of Anti-pneumococcus serum in Collett (2003, pages 6-7). This example is based on the assay taken from Smith (1932), who described a study of the protective effect of a particular serum, 'Serum number 32', on pneumococcus, the bacterium associated with the occurrence of pneumonia. Each of 40 mice was injected with a combination of an infecting dose of a culture of pneumococci and one of five doses of the anti-pneumococcus serum. For all mice that died during the seven-day period following inoculation, a blood smear taken from the heart was examined to determine

Dose of serum	Number of deaths out of 40
0.0028	35
0.0056	21
0.0112	9
0.0225	6
0.0450	1

Table 6.1: Number of deaths to different doses of serum

whether pneumococci were present or absent. Mice that still lived on the seventh day were regarded as survivors and not further examined. The dichotomous response variable is therefore death from pneumonia within seven days of inoculation. The numbers of mice succumbing to infection out of 40 exposed to each of five doses of the serum, measured in cc, are given in Table 6.1.

Obviously, a simple logistic regression model is used to fit the data in order to find the relationship between the probability of deaths and dose of serum. A simultaneous confidence band can be constructed. Here we choose [0, 0.0450] as the restricted interval for the dose of serum.

With the same notations as before, we have

$$\hat{\beta}_0 = 1.2179, \quad \hat{\beta}_1 = -146.6927,$$

$$\mathcal{J}^{-1} = \begin{pmatrix} 0.0858 & -6.1499 \\ -6.1499 & 695.0059 \end{pmatrix}.$$

Also, we have the critical values 2.4304 for a WB band and 2.3700 for a Type 4 band respectively at 95% confidence level. Therefore, two simultaneous confidence bands for the probability of deaths can be constructed accordingly. Both bands are given in Figure 6.2.

6.2.2 For a multiple logistic regression

Method of Piegorsch and Casella (1988)

Piegorsch and Casella (1988) proposed a method of constructing confidence bands for a multiple logistic regression with predictor variables restricted in a rectangular region given by (6.17).

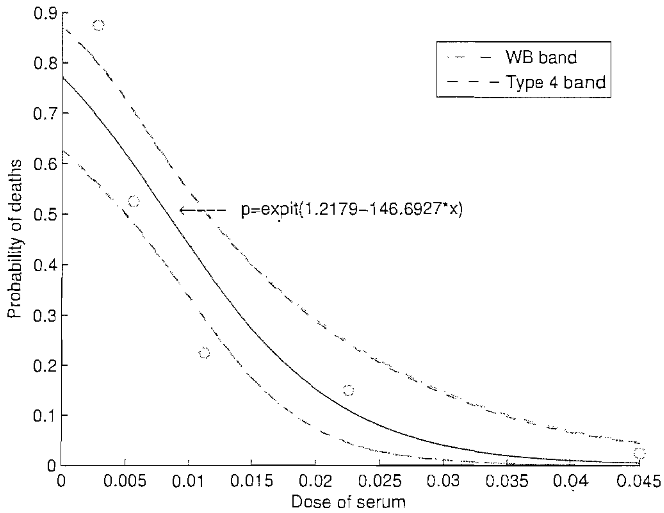

Figure 6.2: 95%-level confidence bands for probability of deaths

The key idea is to embed the rectangular region (6.17) into an ellipsoidal restricted region of predictor variables described in Casella and Strawderman (1980), and then to apply Casella and Strawderman (1980)'s Table 1 to obtain a conservative critical value. In particular, the ellipsoidal region is centered at the means of the predictor variables. If the rectangular region is not centered at the mean point, the critical value obtained from Casella and Strawderman (1980)'s results can be extremely conservative.

Consider, for example, the quantal data in Table 1 of Piegorsch and Casella (1988). Table 5 gives values of c^2 for 95% confidence level based on their method . Note that this is a one-dimensional example, and all three restricted intervals in Table 5 are asymmetric about the mean of the only predictor variable, which is 1.4862. We used the method of Wynn and Bloomfield (1971) to calculate the critical values for the confidence bands with these three restricted intervals respectively and then compare the squared values of them with those of Piegorsch and Casella (1988). All the values are given in Table 6.2.

Restricted interval	c^2	c_wb^2
[-1.3, 2.0]	5.98	5.4943
[-1.3, 0.8]	5.17	4.8638
[-1.3, -0.2]	4.71	4.2737

*Values in first two columns taken from Piegorsch and Casella (1988).

Table 6.2: Squared critical values for 95% confidence level

From the table, the squared critical values based on the method of Piegorsch and Casella (1988) are, respectively, 8.8%, 6.3% and 10.2% larger than those based on the method of Wynn and Bloomfield (1971). Consequently, the method of Piegorsch and Casella (1988) is not considered further in this chapter.

Simulation-based method

Liu, Jamshidian, Zhang and Donnelly (2005) construct simultaneous confidence bands for a multiple linear regression over a rectangular restricted predictor space based on simulation. We apply the method to the logistic regression case.

For a logistic regression with at least one predictor variables, $\hat{\beta}$ is approximately normally distributed with mean β and estimated asymptotic covariance matrix J^{-1}. Then the coverage probability of a confidence band for $\mathbf{x}^T\beta$ over $\mathbf{x} \in \mathcal{X}$ is given by $\mathrm{P}\{T \leq c\}$, where

$$T = \sup_{\mathbf{x}\in\mathcal{X}} \frac{|\mathbf{x}^T(\hat{\beta}-\beta)|}{\sqrt{\mathbf{x}^T J^{-1}\mathbf{x}}}, \tag{6.24}$$

c is a critical value, and $\mathcal{X}$ is a rectangular region of the form given by (6.17). Assume there is a $p \times p$ non-singular matrix P such that $P^T P = J^{-1}$. Then (6.24) can be further written as

$$\begin{aligned} T &= \sup_{\mathbf{x}\in\mathcal{X}} \frac{|(P\mathbf{x})^T(P^{-1})^T(\hat{\beta}-\beta)|}{\sqrt{(P\mathbf{x})^T(P\mathbf{x})}} \\ &= \sup_{\mathbf{x}\in\mathcal{X}} \frac{|(P\mathbf{x})^T\mathbf{N}|}{\|P\mathbf{x}\|}, \end{aligned} \tag{6.25}$$

where $\mathbf{N} = (P^{-1})^T(\hat{\beta} - \beta)$ has a multivariate standard normal distribution.

The key of the simulation-based method is to generate an $\mathbf{N}$ so as to find a T via (6.25); repeat this process R times, and set the $[(1-\alpha)R]$th largest value of T, $\hat{c}$, as an approximate of the critical value c. For each simulation, T is obtained by solving an optimization problem. Details can be found in Liu, Jamshidian, Zhang and Donnelly (2005). After $\hat{c}$ is obtained, a confidence band can be constructed for the logistic response as before.

A numerical example

This is another example from Collett (2003, pages 8-9). The erythrocyte sedimentation rate (ESR) is the rate at which red blood cells (erythrocytes) settle out of suspension in blood plasma, when measured under standard conditions. The ESR increases if the level of certain proteins in the blood plasma rise, such as in rheumatic diseases, chronic infections and malignant diseases; this makes the determination of the ESR one of the most commonly used screening tests performed on samples of blood. One aspect of a study carried out by the Institute of Medical Research, Kuala Lumpur, Malaysia, was to examine the extent to which the ESR is related to two plasma proteins, fibrinogen and γ-globulin, both measured in gm/l, for a sample of 32 individuals. The ESR for a 'healthy' individual should be less than 20 mm/hr and since the absolute value of the ESR is relatively unimportant, the response variable used here will denote whether this is the case. A response of zero will signify a healthy individual (ESR$<$20), while a response of unity will refer to an unhealthy individual (ESR$\geq$20). The original data were presented in Collett and Jemain (1985) and are relisted in Table 6.3.

In this case, a bivariate logistic regression model is applied to obtain the relationship between the probability of an ESR reading greater than 20 mm/hr and the levels of two plasma proteins. When construct a simultaneous confidence band, we set an restricted interval for each predictor variable formed by the smallest and largest values of the observations. Specifically, they are [2.09, 5.06] and [28, 46] for Fibrinogen and γ-globulin respectively.

Individual	Fibrinogen	r-globulin	Response
1	2.52	38	0
2	2.56	31	0
3	2.19	33	0
4	2.18	31	0
5	3.41	37	0
6	2.46	36	0
7	3.22	38	0
8	2.21	37	0
9	3.15	39	0
10	2.60	41	0
11	2.29	36	0
12	2.35	29	0
13	5.06	37	1
14	3.34	32	1
15	2.38	37	1
16	3.15	36	0
17	3.53	46	1
18	2.68	34	0
19	2.60	38	0
20	2.23	37	0
21	2.88	30	0
22	2.65	46	0
23	2.09	44	1
24	2.28	36	0
25	2.67	39	0
26	2.29	31	0
27	2.15	31	0
28	2.54	28	0
29	3.93	32	1
30	3.34	30	0
31	2.99	36	0
32	3.32	35	0

Table 6.3: The levels of two plasma proteins and the value of a binary response that denotes whether ESR$\geq$20 for each individual

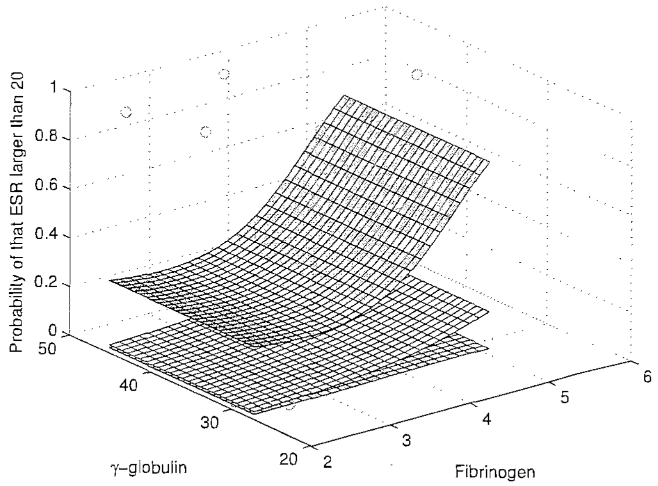

Figure 6.3: 90%-level confidence band for probability of that ESR larger than 20

With the same notations as before, we have

$$\hat{\beta} = \begin{pmatrix} -12.7921 \\ 1.9104 \\ 0.1558 \end{pmatrix} \text{ and } J^{-1} = \begin{pmatrix} 33.5985 & -3.6718 & -0.5987 \\ -3.6718 & 0.9428 & 0.0224 \\ -0.5987 & 0.0224 & 0.0143 \end{pmatrix}.$$

Also, we have the critical values 2.1291 and 2.4118 based on the simulation method of 100,000 simulations for 90% and 95% confidence level respectively. Two simultaneous confidence bands can be constructed then. They are plotted in Figures 6.3 and 6.4.

6.3 Simulations

All these methods of constructing simultaneous confidence bands for a logistic regression is based on the large sample asymptotic normality of $\hat{\beta}$. So the bands constructed have an approximate $1-\alpha$ confidence level. It is therefore of interest to simulate the coverage probabilities of the bands to check how close they are to the nominal level, and what factors affect the accuracy.

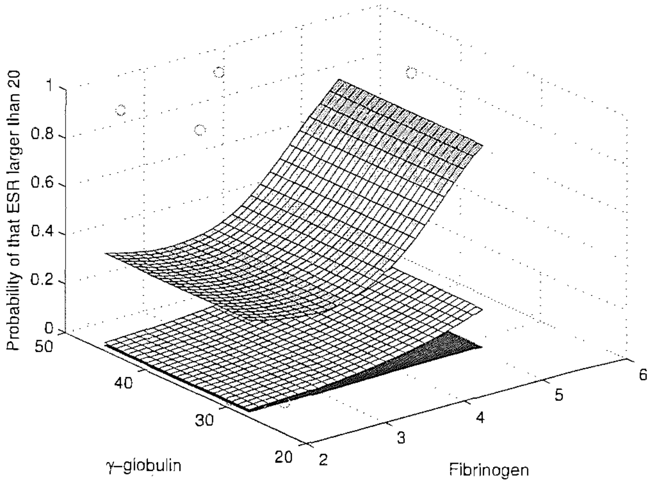

Figure 6.4: 95%-level confidence band for probability of that ESR larger than 20

We have carried out simulation studies for the one-dimensional and two-dimensional cases, respectively.

6.3.1 For one-dimensional case

In this subsection, we compare the confidence band based on the method of Wynn and Bloomfield (1971) with the Type 4 band recommended in Sun, Loader and McCormick (2000). We call them WB band and Type 4 band for simplicity.

With consistent notations, the specific procedure is as follow:

Step 1 Given a set of m values of the only predictor variable x, $x_1, \dots, x_m$, together with a pair of true regression coefficients β_0 and β_1, we obtain the probabilities of the logistic response based on the true model via

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}, i = 1, \dots, m. \tag{6.26}$$

Step 2 Generate the logistic response y_i, corresponding to x_i, which has a binomial distribution with parameters $\pi(x_i)$ and n_i, where n_i is the sub-sample size.

Step 3 Estimate β_0 and β_1 based on $(x_i, y_i, n_i), i = 1, \ldots, m$ to obtain $\hat{\beta}_0$ and $\hat{\beta}_1$, then calculate J^{-1} accordingly.

Step 4 Construct a simultaneous confidence band. WB band is of the form

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm c_1 \sqrt{(1, x) J^{-1} (1, x)^T} \quad x \in [a, b], \tag{6.27}$$

where c_1 can be calculated using (6.21) and (6.22) for a given nominal confidence level cl. Type 4 band is of the form

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm (c_2 - |\hat{r}_p|) \sqrt{(1, x) J^{-1} (1, x)^T} \quad x \in [a, b], \tag{6.28}$$

where $(c_2 - |\hat{r}_p|)$ as a whole can be obtained directly by using the software `parfit` mentioned in Sun, Loader and McCormick (2000).

Step 5 Check whether the true function $\beta_0 + \beta_1 x$ is completely in the band for all x's within the restricted interval $[a, b]$.

We repeat Step 1 to Step 5 $nsim$ times. Out of the $nsim$ simulations, the proportion of times that the confidence band includes the true regression model is taken as an approximation of the true coverage probability. It is clear that this simulated coverage probability depends on $x_i, n_i, \beta_0, \beta_1, [a, b], cl, nsim$, $i = 1, \ldots, m$.

Now, we turn to the design of these common factors so that various combinations can be obtained to make the comparison as general as possible. First, we choose five design points for the only predictor variable, that is $m = 5$, which seems reasonable in real problems. Furthermore, we choose four different types of five design points on the design interval $[-1, 1]$. The first type of design is to set the design points equally spaced throughout the interval $[-1, 1]$. The second type corresponds to the design points near the center of the design interval. The third type corresponds to the design

Structure	Design points	Restricted interval	Θ^*
P1	[-1, -0.5, 0, 0.5, 1]	[-0.5, 0.5]	0.6155
P2	*Equally-spaced*	[-2, 2]	1.2310
P3	[-0.06, -0.05, 0.02, 0.04, 0.09]	[-0.5, 0.5]	1.4484
P4	*Centred*	[-2, 2]	1.5400
P5	[0.89, 0.92, 0.95, 0.98, 1]	[-0.5, 0.5]	0.0305
P6	*One-ended*	[-2, 2]	1.5452
P7	[-1, -0.91, 0.94, 0.98, 1]	[-0.5, 0.5]	0.4721
P8	*Two-ended*	[-2, 2]	1.1260

Table 6.4: Designs for predictor variable and restricted interval

Structure	Sample size	Total
N1	[10, 10, 10, 10, 10]	N = 50
N2	[22, 35, 58, 46, 39]	N = 200

Table 6.5: Design for total sample size

points concentrated around the upper bound of the design interval. And the last type corresponds to the design points located at the two ends of the interval $[-1, 1]$. Second, two restricted intervals are chosen, one of which is short and the other is long. These choices of the design points and the restricted intervals provide various values of the angle θ^* in (6.21) to give various critical values. In fact, based on our designs, values of θ^* varies in the range $[0, \pi]$ from small (0.0305) to large (1.5452). Third, we choose two sample sizes $N = \sum_{i=1}^{5} n_i$ of 50 and 200 to check its effect on the simulated coverage probability. Finally, we choose eight pairs of β_0 and β_1 so that the straight line $\beta_0 + \beta_1 x$ has various slopes and intercepts. The designs are contained in Tables 6.4-6.6.

Structure	β0	β1
Q1	0.75	0.5
Q2	2.55	1.7
Q3	-0.39	-0.26
Q4	-1.5	-1
Q5	-0.75	0.5
Q6	-2.55	1.7
Q7	0.39	-0.26
Q8	1.5	-1

Table 6.6: Designs for true regression coefficients

The simulated coverage probabilities of WB bands are calculated by the programmes running on MATLAB 7 platform, while those of Type 4 bands are obtained by using `parfit` on S-plus 6.2 platform. As it is not clear how to change the default confidence level 95% in `parfit`, confidence level 95% is used in our comparison. To reduce the simulation error, we set $nsim$ equal to 10,000. Based on the above designs, simulated coverage probabilities are obtained and listed in Tables 6.7-6.10. Note that, in some cases, the simulated coverage probability can not be worked out. If this is the case for both bands based on the same designs, it is because the maximum likelihood estimates of (β_0, β_1) can not be found within the pre-specified 30 iterations. The case that only Type 4 band can not find the simulated coverage probability is due to the fact that the corrected critical value $(c - |\hat{r}_p|)$ can not be found using `parfit`. In this case, a sentence "`warning: compparcomp: perfect fit`" was displayed.

From the results, some conclusions can be drawn. First, the simulated coverage probabilities of both bands are often larger than 95%. Second, when the sample size $N = 50$, both bands can be quite conservative with the simulated coverage probabilities being around 97%. When $N = 200$, the simulated coverage probabilities of WB bands are very close to 95% except few cases, whereas Type 4 bands may still be quite conservative or liberal. Third, the corrected critical value of Type 4 band may not be found for small sample size.

Also, it is motivated to compare the widths of the WB bands and Type 4 bands. Note that there are 128 design structures in our comparison, such as P1,N1,Q1 and P5,N2,Q8. For each design structure, we calculate 100 simulated critical values for each band. Therefore, we have totally 12,800 critical values for each band. Ignore the cases that one or both bands can not find the critical value. The left 10,476 cases are viewed as being valid. Then it is found that the proportion of the cases that the critical value of WB band is smaller than that of Type 4 band, out of the valid cases, is 68.36%.

Consequently, WB band seems to be better than Type 4 band generally,

Structure			scp_wb	scp_t4
P1	N1	Q1	0.9702	0.9652
		Q2	Na	Na
		Q3	0.9730	0.9706
		Q4	0.9539	0.9417
		Q5	0.9720	0.9652
		Q6	Na	Na
		Q7	0.9729	0.9715
		Q8	0.9529	0.9409
	N2	Q1	0.9510	0.9582
		Q2	0.9431	0.9412
		Q3	0.9529	0.9575
		Q4	0.9568	0.9587
		Q5	0.9521	0.9570
		Q6	0.9238	0.9320
		Q7	0.9497	0.9569
		Q8	0.9511	0.9559
P2	N1	Q1	0.9728	0.9832
		Q2	Na	Na
		Q3	0.9684	0.9801
		Q4	0.9575	0.9672
		Q5	0.9744	0.9807
		Q6	Na	Na
		Q7	0.9681	0.9805
		Q8	0.9612	0.9685
	N2	Q1	0.9484	0.9615
		Q2	0.9446	0.9537
		Q3	0.9523	0.9648
		Q4	0.9558	0.9673
		Q5	0.9510	0.9672
		Q6	0.9358	0.9536
		Q7	0.9533	0.9663
		Q8	0.9588	0.9683

Table 6.7: Simulated coverage probabilities for 95% confidence level

especially when the total sample size is large.

Structure			scp_wb	scp_t4
P3	N1	Q1	0.9730	Na
		Q2	Na	Na
		Q3	0.9696	Na
		Q4	0.9879	Na
		Q5	0.9717	Na
		Q6	Na	Na
		Q7	0.9703	Na
		Q8	0.9878	Na
	N2	Q1	0.9490	0.9666
		Q2	0.9753	0.9782
		Q3	0.9470	Na
		Q4	0.9565	0.9702
		Q5	0.9516	0.9666
		Q6	0.9659	0.9735
		Q7	0.9518	0.9695
		Q8	0.9620	0.9722
P4	N1	Q1	0.9739	Na
		Q2	Na	Na
		Q3	0.9648	Na
		Q4	0.9833	Na
		Q5	0.9736	Na
		Q6	Na	Na
		Q7	0.9638	Na
		Q8	0.9889	Na
	N2	Q1	0.9492	0.9675
		Q2	0.9686	0.9803
		Q3	0.9582	Na
		Q4	0.9515	0.9732
		Q5	0.9501	0.9698
		Q6	0.9630	0.9787
		Q7	0.9523	0.9703
		Q8	0.9551	0.9733

Table 6.8: Simulated coverage probabilities for 95% confidence level

Structure			scp_wb	scp_t4
P5	N1	Q1	0.9859	Na
		Q2	Na	Na
		Q3	0.9744	0.9343
		Q4	Na	Na
		Q5	0.9628	Na
		Q6	0.9784	Na
		Q7	0.9633	Na
		Q8	0.9665	Na
	N2	Q1	0.9532	0.9369
		Q2	Na	Na
		Q3	0.9466	0.9245
		Q4	0.9688	0.9588
		Q5	0.9489	0.9292
		Q6	0.9550	0.9344
		Q7	0.9490	0.9305
		Q8	0.9504	0.9259
P6	N1	Q1	0.9627	Na
		Q2	Na	Na
		Q3	0.9677	0.9809
		Q4	Na	Na
		Q5	0.9652	Na
		Q6	0.9743	Na
		Q7	0.9641	Na
		Q8	0.9705	Na
	N2	Q1	0.9519	0.9689
		Q2	Na	Na
		Q3	0.9523	0.9718
		Q4	0.9660	0.9798
		Q5	0.9490	0.9702
		Q6	0.9536	0.9699
		Q7	0.9493	0.9688
		Q8	0.9502	0.9709

Table 6.9: Simulated coverage probabilities for 95% confidence level

Structure			scp_wb	scp_t4
P7	N1	Q1	0.9699	0.9646
		Q2	Na	Na
		Q3	0.9688	0.9651
		Q4	0.9443	0.9341
		Q5	0.9678	0.9561
		Q6	0.8803	Na
		Q7	0.9751	0.9620
		Q8	0.9574	0.9205
	N2	Q1	0.9539	0.9548
		Q2	0.9129	0.8814
		Q3	0.9542	0.9542
		Q4	0.9584	0.9557
		Q5	0.9569	Na
		Q6	0.9150	0.8630
		Q7	0.9542	0.9540
		Q8	0.9589	0.9492
P8	N1	Q1	0.9693	0.9712
		Q2	Na	Na
		Q3	0.9683	0.9726
		Q4	0.9464	0.9345
		Q5	0.9701	0.9662
		Q6	0.9027	Na
		Q7	0.9707	0.9739
		Q8	0.9566	0.9325
	N2	Q1	0.9551	0.9680
		Q2	0.9344	0.9309
		Q3	0.9539	0.9702
		Q4	0.9579	0.9671
		Q5	0.9597	Na
		Q6	0.9318	0.9256
		Q7	0.9501	0.9652
		Q8	0.9569	0.9653

Table 6.10: Simulated coverage probabilities for 95% confidence level

Structure	Design points	Types
P1	x1=[0.35, 0.39, 0.48, 0.52, 0.57, 0.61]	*Centred*
Centred	x2=[0.37, 0.41, 0.5, 0.53, 0.56, 0.58]	*Centred*
P2	x1=[0.85, 0.88, 0.92, 0.94, 0.98, 1]	*Right-ended*
One-cornered	x2=[0.03, 0.07, 0.11, 0.14, 0.16, 0.19]	*Left-ended*
P3	x1=[0.05, 0.08, 0.1, 0.92, 0.95, 0.98]	*Two-ended*
Two-cornered	x2=[0.02, 0.04, 0.07, 0.91, 0.96, 0.99]	*Two-ended*
P4	x1=[0.2, 0.4, 0.7, 0.5, 0.8, 0.9]	*Non-equally-spaced*
Dispersed	x2=[0.15, 0.3, 0.45, 0.6, 0.75, 0.9]	*Equally-spaced*

Table 6.11: Design points for two predictor variables

Structure	Restricted intervals	Type
R1	[0, 2] , [0, 2]	*Long - Long*
R2	[0.5, 1] , [0.5, 1]	*Short - Short*
R3	[0.5, 1] , [0, 2]	*Short - Long*

Table 6.12: Designs for restricted intervals of predictor variables

6.3.2 For two-dimensional case

For the two-dimensional case, we find the simulated coverage probabilities of the confidence bands constructed based on the simulation method of Liu, Jamshidian, Zhang and Donnelly (2005), and compare the results with the nominal confidence level.

The procedure is very similar to that in the one-dimensional case. The only difference is to change the number of predictor variables from 1 to 2 and the consequential changes to the regression coefficients, the restricted region, and the critical values.

Specifically, we choose four different designs P1,P2,P3,P4 in the predictor space, three pairs of restricted intervals R1,R2,R3, two levels of the total sample size N1,N2, and eight sets of the true regression coefficients Q1-Q8. Details are clearly shown in Tables 6.11-6.14.

Structure	Sample size	Total
N1	[10, 10, 10, 10, 10, 10]	N = 60
N2	[16, 25, 29, 34, 46, 50]	N = 200

Table 6.13: Designs for total sample size

Structure	β0	β1	β2
Q1	-0.7	1.7	1.7
Q2	0.25	0.5	1
Q3	1.25	0.5	-1
Q4	1.5	-0.5	-0.5
Q5	-2.7	1.7	1.7
Q6	-1.75	0.5	1
Q7	-0.75	0.5	-1
Q8	-0.5	-0.5	-0.5

Table 6.14: Designs for true regression coefficients

90% and 95% confidence levels are chosen in this simulation study. Note that this time when we construct confidence bands, the critical values come from the simulation-based method. We set the number of simulations equal to 5,000 for the calculation of the critical value, and the number of simulations equal to 10,000 for the calculation of the coverage probability. We consider this setting of the number of simulations as Type 1 setting. Alternatively, we may set 10,000 simulations for the critical value's calculation and 5,000 simulations for the coverage probability's calculation, which is considered as Type 2 setting. We have tried ten specific cases based on both settings, among which five are for the small sample size and the other five are for the large sample size. By comparing the resulting simulated coverage probabilities, it is found that the difference between the simulated coverage probabilities for the Type 1 and Type 2 settings is at the third decimal place for all ten chosen cases. So we reasonably believe that using either one may not influence our conclusions. Since it will take long time to do simulations for both settings, we just choose Type 1 setting here. Results are given in Tables 6.15-6.18.

From these results, it can be concluded that when $N = 60$ the confidence bands constructed based on the simulation method are much conservative with the simulated coverage probabilities generally larger than 93% for 90% confidence level and 97% for 95% level. When $N = 200$ the simulated coverage probabilities are pretty close to the nominal confidence levels, sometimes larger and sometimes smaller. Consequently, we reasonably believe this kind of confidence bands are good enough when N is large.

Structure combinations			scp 90%	95%
P1, R1	N1	Q1	0.9404	0.9775
		Q2	0.9418	0.9798
		Q3	0.9395	0.9763
		Q4	0.9428	0.9785
		Q5	0.9454	0.9776
		Q6	0.9417	0.9780
		Q7	0.9437	0.9785
		Q8	0.9360	0.9804
	N2	Q1	0.8945	0.9445
		Q2	0.9077	0.9525
		Q3	0.9033	0.9508
		Q4	0.8956	0.9501
		Q5	0.9022	0.9554
		Q6	0.9021	0.9513
		Q7	0.8977	0.9485
		Q8	0.8996	0.9476
P1, R2	N1	Q1	0.9401	0.9766
		Q2	0.9399	0.9763
		Q3	0.9455	0.9777
		Q4	0.9460	0.9799
		Q5	0.9382	0.9757
		Q6	0.9409	0.9790
		Q7	0.9447	0.9794
		Q8	0.9420	0.9767
	N2	Q1	0.8943	0.9456
		Q2	0.9057	0.9549
		Q3	0.9012	0.9456
		Q4	0.9028	0.9525
		Q5	0.9138	0.9587
		Q6	0.9034	0.9514
		Q7	0.8967	0.9519
		Q8	0.8934	0.9499
P1, R3	N1	Q1	0.9349	0.9736
		Q2	0.9329	0.9737
		Q3	0.9305	0.9744
		Q4	0.9332	0.9746
		Q5	0.9275	0.9749
		Q6	0.9291	0.9752
		Q7	0.9316	0.9748
		Q8	0.9341	0.9722
	N2	Q1	0.9014	0.9543
		Q2	0.9050	0.9523
		Q3	0.8916	0.9490
		Q4	0.9049	0.9553
		Q5	0.9005	0.9493
		Q6	0.9026	0.9496
		Q7	0.8890	0.9471
		Q8	0.8969	0.9468

Table 6.15: Simulated coverage probabilities for two-dimensional case

Structure combinations			scp	
			90%	95%
P2, R1	N1	Q1	0.9372	0.9801
		Q2	0.9262	0.9694
		Q3	0.9559	0.9870
		Q4	0.9360	0.9771
		Q5	0.9268	0.9735
		Q6	0.9376	0.9791
		Q7	0.9158	0.9588
		Q8	0.9390	0.9758
	N2	Q1	0.8990	0.9535
		Q2	0.9065	0.9526
		Q3	0.9046	0.9551
		Q4	0.6889	0.9467
		Q5	0.9068	0.9526
		Q6	0.8945	0.9500
		Q7	0.8952	0.9502
		Q8	0.8887	0.9465
P2, R2	N1	Q1	0.9374	0.9772
		Q2	0.9294	0.9667
		Q3	0.9614	0.9895
		Q4	0.9264	0.9708
		Q5	0.9197	0.9745
		Q6	0.9378	0.9749
		Q7	0.9064	0.9622
		Q8	0.9381	0.9754
	N2	Q1	0.8975	0.9439
		Q2	0.8976	0.9445
		Q3	0.9054	0.9550
		Q4	0.8962	0.9516
		Q5	0.8975	0.9466
		Q6	0.8953	0.9464
		Q7	0.8998	0.9505
		Q8	0.8899	0.9430
P2, R3	N1	Q1	0.9370	0.9758
		Q2	0.9207	0.9739
		Q3	0.9587	0.9888
		Q4	0.9268	0.9726
		Q5	0.9176	0.9664
		Q6	0.9393	0.9782
		Q7	0.9131	0.9588
		Q8	0.9374	0.9738
	N2	Q1	0.8968	0.9460
		Q2	0.8947	0.9413
		Q3	0.9010	0.9467
		Q4	0.8958	0.9443
		Q5	0.8887	0.9479
		Q6	0.8995	0.9494
		Q7	0.9027	0.9515
		Q8	0.8971	0.9438

Table 6.16: Simulated coverage probabilities for two-dimensional case

Structure combinations			scp 90%	95%
P3, R1	N1	Q1	0.9249	0.9687
		Q2	0.9447	0.9747
		Q3	0.9400	0.9739
		Q4	0.9325	0.9779
		Q5	0.9200	0.9649
		Q6	0.9285	0.9746
		Q7	0.9421	0.9760
		Q8	0.9454	0.9796
	N2	Q1	0.9092	0.9521
		Q2	0.8906	0.9464
		Q3	0.9042	0.9553
		Q4	0.9045	0.9566
		Q5	0.9049	0.9560
		Q6	0.9004	0.9496
		Q7	0.8912	0.9474
		Q8	0.8979	0.9465
P3, R2	N1	Q1	0.9067	0.9584
		Q2	0.9378	0.9731
		Q3	0.9393	0.9739
		Q4	0.9365	0.9750
		Q5	0.9338	0.9775
		Q6	0.9282	0.9704
		Q7	0.9414	0.9742
		Q8	0.9391	0.9706
	N2	Q1	0.9074	0.9627
		Q2	0.8929	0.9463
		Q3	0.9001	0.9505
		Q4	0.8969	0.9521
		Q5	0.9029	0.9552
		Q6	0.8963	0.9540
		Q7	0.8956	0.9486
		Q8	0.8971	0.9472
P3, R3	N1	Q1	0.9504	0.9813
		Q2	0.9442	0.9827
		Q3	0.9286	0.9763
		Q4	0.9229	0.9689
		Q5	0.9127	0.9670
		Q6	0.9220	0.9688
		Q7	0.9428	0.9748
		Q8	0.9471	0.9809
	N2	Q1	0.9045	0.9540
		Q2	0.8918	0.9446
		Q3	0.8942	0.9475
		Q4	0.8963	0.9503
		Q5	0.8994	0.9494
		Q6	0.8997	0.9505
		Q7	0.8930	0.9468
		Q8	0.9026	0.9521

Table 6.17: Simulated coverage probabilities for two-dimensional case

Structure combinations			scp 90%	scp 95%
P4, R1	N1	Q1	0.9350	0.9691
		Q2	0.9407	0.9796
		Q3	0.9444	0.9746
		Q4	0.9389	0.9754
		Q5	0.9276	0.9702
		Q6	0.9412	0.9734
		Q7	0.9457	0.9798
		Q8	0.9467	0.9776
	N2	Q1	0.8913	0.9482
		Q2	0.8874	0.9432
		Q3	0.8978	0.9499
		Q4	0.9031	0.9511
		Q5	0.9046	0.9487
		Q6	0.9046	0.9512
		Q7	0.9059	0.9514
		Q8	0.8912	0.9443
P4, R2	N1	Q1	0.9191	0.9628
		Q2	0.9337	0.9728
		Q3	0.9431	0.9776
		Q4	0.9363	0.9778
		Q5	0.9306	0.9707
		Q6	0.9378	0.9762
		Q7	0.9428	0.9737
		Q8	0.9415	0.9713
	N2	Q1	0.9016	0.9459
		Q2	0.8977	0.9488
		Q3	0.8990	0.9515
		Q4	0.8943	0.9554
		Q5	0.9094	0.9508
		Q6	0.9007	0.9505
		Q7	0.8961	0.9495
		Q8	0.8951	0.9458
P4, R3	N1	Q1	0.9451	0.9764
		Q2	0.9375	0.9757
		Q3	0.9386	0.9718
		Q4	0.9342	0.9736
		Q5	0.9246	0.9740
		Q6	0.9352	0.9706
		Q7	0.9394	0.9740
		Q8	0.9413	0.9740
	N2	Q1	0.9030	0.9534
		Q2	0.9043	0.9533
		Q3	0.8963	0.9551
		Q4	0.9035	0.9517
		Q5	0.9061	0.9516
		Q6	0.8998	0.9499
		Q7	0.8943	0.9478
		Q8	0.8968	0.9521

Table 6.18: Simulated coverage probabilities for two-dimensional case

6.3.3 Conclusions

From the simulation results for both one-dimensional and two-dimensional cases, it is clear that the total sample size N plays a central role. When N is small, the confidence bands tends to be conservative. But for a sufficiently large N, the simulated coverage probabilities are often very close to the nominal confidence levels. This observation agrees with the large sample theory which is the base of the construction of simultaneous confidence bands for generalized linear models.

Chapter 7

Conclusions and future work

7.1 Conclusions

This thesis considers the construction of simultaneous confidence bands for a classical normal-error linear regression model and a general linear logistic regression model with a binary response variable. From the work in the last few chapters, some main conclusions can be drawn.

7.1.1 For linear regression

For linear regression, the confidence bands, centered by the estimated mean responses $\mathbf{x}^{\mathbf{T}}\hat{\beta}$ and with the band width proportional to the standard error of $\mathbf{x}^{\mathbf{T}}\hat{\beta}$, are of interest. This type of confidence bands are of hyperbolic shape following Scheffé (1953)'s procedure, and are more popular than the bands of other shapes. Also, it is important to impose a constraint on each predictor variable so that the confidence bands constructed over the obtained restricted region are not unnecessarily wide when we deal with a real problem. Therefore, constructing exact confidence bands over different restricted regions becomes the central task. Two most frequently mentioned regions are the ellipsoidal region that centered at the point of the means of the predictor variables, and the rectangular region that is formed by imposing an interval constraint on each predictor variable.

Over an ellipsoidal region, this thesis summarizes three methods to construct both one-sided and two-sided exact simultaneous confidence bands for a linear regression model. These three methods are: the method following the idea of Bohrer (1973), the algebraical method, and the tubular neighborhood method. They start from the same point and are proven to have the equivalent computational formulae for calculation of critical values. Furthermore, it is found that the first method has a relatively simple computational formula for both simple and multiple linear regression cases. In addition, for some special cases, these three methods agree with some other well-known methods in statistical literatures. For instance, the first method of constructing one-sided and two-sided bands for a simple linear regression was considered by Bohrer and Francis (1972) and Wynn and Bloomfield (1971) respectively. The algebraical method of constructing two-sided bands for a multiple linear regression was considered by Casella and Strawderman (1980). Moreover, the idea of the tubular neighborhood method also appeared in Naiman (1986), Sun and Loader (1994).

To construct simultaneous confidence bands for a regression model over a rectangular region, several methods are available. Among these methods, Naiman (1986) produced a conservative confidence band for one-dimensional regression models, and his idea may be applied to the high dimensional cases but no explicit computational formula was given. The approximate method of Sun and Loader (1994) considered an approximation to the tube formula. The simulation-based method of Liu, Wynn and Hayter (2005) and Liu, Jamshidian, Zhang and Donnelly (2005) for polynomial regression and multiple linear regression respectively used Monte Carlo simulation. This thesis compares these methods in terms of critical values for simple linear regression, polynomial regression and bivariate linear regression respectively. From the simulation results, several conclusions can be drawn. The simulation-based method of Liu *et al.* (2005) can compute critical values almost as accurate as the exact method for a simple linear regression. It is better than the conservative method of Naiman (1986) and the approximate method of Sun and Loader (1994) for polynomial and bivariate linear regressions. Naiman's method is quite good in the sense that it is a conservative method but its critical values are actually not too conservative. The approximate method, comparatively speaking, is bad but not seriously. These conclusions may be useful for someone who wants to construct simultaneous confidence bands for data analysis.

7.1.2 For logistic regression

For logistic regression, the base of constructing simultaneous confidence bands is the asymptotic normal distribution of the estimator. Hence, this thesis first illustrates a way of finding the asymptotic normality of the maximum likelihood estimator of the parameters of interest following the idea presented in Sen and Singer (1993).

Existing literatures on the construction of confidence bands for a logistic regression model are very limited. Methods of Brand, Pinnock and Jackson (1973) and Hauck (1983) construct confidence bands over the whole predictor space for simple and multiple regression cases respectively. Over a rectangular restricted region, we consider the Type 4 band of Sun, Loader and McCormick (2000) specially for the one-dimensional logistic regression. The method of Piegorsch and Casella (1988) for the construction of confidence bands for a multiple logistic regression is found not to be recommendable. This thesis considers two new methods following the ideas of Wynn and Bloomfield (1971) for simple regression and Liu, Jamshidian, Zhang and Donnelly (2005) for multiple regression. The confidence bands produced by these two methods are named WB band and Simulation-based band accordingly.

To assess the performance of these confidence bands, this thesis provides simulation studies for both one-dimensional and two-dimensional cases. From the simulated results, some useful conclusions can be drawn. The bands obtained based on large-size samples are better than those with small-size samples in the sense that the simulated coverage probabilities of the bands are clearly more closer to the nominal confidence levels for larger sample sizes. For one-dimensional case, WB band seems to be better than Type 4 band recommended by Sun, Loader and McCormick (2000), and it often performs very well when the sample size N is 200. For two-dimensional case, the Simulation-based bands are quite conservative when $N = 60$, and its simulated coverage probabilities are very close to the nominal confidence levels when $N = 200$.

7.2 Future work

This thesis considers constructing simultaneous confidence bands for only one regression model. The construction of confidence bands for two or more than two regression models may be of interest in the future work. Also, we only focus on the construction of confidence bands for a logistic regression with binary data in the thesis. We may consider constructing confidence bands for the ordinal logistic regression and the multinomial logistic regression. Moreover, we may think about other classes of regression models in the family of generalized linear models. Since all members of the generalized linear models share the large sample asymptotic property, the methods of constructing confidence bands should be similar.

Appendix A

Codes for computing the critical value and simulated coverage probability

This appendix provides the codes for computing the critical values using various methods for simple linear regression, polynomial regression and bivariate linear regression, and for computing the simulated coverage probabilities for the one-dimensional and two-dimensional linear logistic regressions. All codes in this appendix are written using MATLAB unless it is particularly specified.

A.1 For computing the critical value for linear regression

A.1.1 Obtaining c, using the exact method for simple linear regression

```
function c_wb=wb_cv(k0,nu,cl)

%%Output
```

```
%c_wb: the critical value of WB method
%%Input
%k0: the angle theta
%nu: the degree of freedom
%cl: the confidence level

tol=0.0001;
N0=10000;
aa=1;bb=10;
faa=@(beta)fcdf((aa^2)/(2*(cos(beta))^2),2,nu);
int_faa=quad(faa,0,pi/2-k0/2);
HA=(2/pi)*(k0/2)*fcdf((aa^2)/2,2,nu)+(2/pi)*int_faa;
i=1;
while i<=N0
    c(i)=aa+(bb-aa)/2;
    f=@(beta)fcdf((c(i)^2)/(2*(cos(beta))^2),2,nu);
    int_f=quad(f,0,pi/2-k0/2);
    HI=(2/pi)*(k0/2)*fcdf((c(i)^2)/2,2,nu)+(2/pi)*int_f;
    if HI-cl==0 | (bb-aa)/2<tol
        c_wb=c(i);break
    end
    i=i+1;
    if HA*(HI-cl)<0
       aa=c(i-1);
       HA=HI;
    else
        bb=c(i-1);
    end
end
return
```

A.1.2 Obtaining c, using the approximate method for simple linear regression

```
function c_app=app_cv(k0,nu,cl)

%%Output
%c_app: the critical value of the approximate method
%%Input
%k0,nu,cl: the same as before

tol=0.0001;
N0=10000;
aa=1;bb=10;
alpha_aa=(k0/pi)*(1+aa^2/nu)^(-nu/2)+2*(1-tcdf(aa,nu));
i=1;
while i<=N0
    c(i)=aa+(bb-aa)/2;
    alpha_I=(k0/pi)*(1+c(i)^2/nu)^(-nu/2)+2*(1-tcdf(c(i),nu));
    if alpha_I-(1-cl)==0 | (bb-aa)/2<tol
        c_app=c(i);break
    end
    i=i+1;
    if alpha_aa*(alpha_I-(1-cl))>0
        aa=c(i-1);
        alpha_aa=alpha_I;
    else
        bb=c(i-1);
    end
end
return
```

A.1.3 Obtaining c, using the simulation-based method for simple linear regression

```
function [c_simu,se]=simu_cv(k0,nu,cl,nsim)

%%Output
%c_simu: the critical value of the simulation-based method
%se: the standard error
%%Input
%k0,nu,cl: the same as before
%nsim: the number of simulations



%Generate sigma^/sigma
S=sqrt(sum(randn(nu,nsim).^2)./nu);

%Compute T
for j=1:nsim
    N=randn(1,2);
    u=N./norm(N);
    if abs(u(1))>=cos(k0/2)
        Q(j)=norm(N);
    elseif ((u(1)>=0) & (u(1)<cos(k0/2)) & (u(2)>sin(k0/2))) |
           ((u(1)<=0) & (u(1)>-cos(k0/2)) & (u(2)<-sin(k0/2)))
        Q(j)=norm(N)*abs(u(1)*cos(k0/2)+u(2)*sin(k0/2));
    else
        Q(j)=norm(N)*abs(u(1)*cos(k0/2)-u(2)*sin(k0/2));
    end
end
T=Q./S;

%Compute the quantile of the simulated values
```

```
T=sort(T);
r=nsim*cl;
c_simu=T(r);

%Compute the standard error of c
d=0.01;
K=(c_simu-T)/d;
g=sum((1/(d*sqrt(2*pi)))*exp(-0.5*(K.^2)))/length(T);
se=sqrt((cl*(1-cl))/(g^2*length(T)));
return
```

A.1.4 Obtaining c, using Naiman's method for polynomial regression

```
function c_naiman=naiman_1d_cc(w,cl,p1,a,b);

%%Output
%c_naiman: the critical value of Naiman's method
%%Input
%w: the design points of the only predictor variable
%cl: the confidence level
%p1: the order of the polynomial regression plus 1
%a: the lower bound of the restricted interval
%b: the upper bound of the restricted interval

n=length(w);
nu=n-p1;
for j=1:p1
    X(:,j)=w'.^(j-1);
end
P=sqrtm(inv(X'*X));
p=fliplr(P);
```

```
for j=1:p1
    sqp(j,:)=conv(p(j,:),p(j,:));
end
sqmolp=sum(sqp);
x=a:(b-a)/1000:b;
for k=1:length(x)
    for j=1:(p1-1)
        xdp(j)=x(k)^(p1-1-j);
    end
    for j=1:((p1-1)*2+1)
        xsqmolp(j)=x(k)^((p1-1)*2+1-j);
    end
    for j=1:p1
        xp(j)=x(k)^(p1-j);
    end
    for j=1:(p1-1)*2
        xdsqmolp(j)=x(k)^((p1-1)*2-j);
    end
    for j=1:p1
        dT(j)=(polyder(p(j,:))*xdp'*sqrt(sqmolp*xsqmolp')-
        p(j,:)*xp'*(1/(2*sqrt(sqmolp*xsqmolp')))*
        (polyder(sqmolp)*xdsqmolp'))/(sqmolp*xsqmolp');
    end
    moldT(k)=norm(dT);
end
for m=1:(length(x)-1)
    T(m)=(moldT(m)+moldT(m+1))/2;
end
k0=((b-a)/1000)*sum(T');

tol=0.0001;
```

```
N0=100000;
aa=1;bb=10;
t=0:(1/(aa*1000)):(1/aa);
for m=1:length(t)
    faa(m)=min((fcdf(2*((aa*t(m))^(-2)-1)/(p1-2),p1-2,2)*
    (k0/pi)+fcdf(((aa*t(m))^(-2)-1)/(p1-1),p1-1,1)),1)*
    fpdf(p1*t(m)^2,nu,p1)*2*p1*t(m);
end
for k=1:(length(t)-1)
    Func_aa(k)=(faa(k)+faa(k+1))/2;
end
alpha_aa=(1/(aa*1000))*sum(Func_aa');
i=1;
while i<=N0
    c(i)=aa+(bb-aa)/2;
    t_I=0:(1/(c(i)*1000)):(1/c(i));
    for m=1:length(t_I)
        f_I(m)=min((fcdf(2*((c(i)*t_I(m))^(-2)-1)/(p1-2),p1-2,2)
        *(k0/pi)+fcdf(((c(i)*t_I(m))^(-2)-1)/(p1-1),p1-1,1)),1)*
        fpdf(p1*t_I(m)^2,nu,p1)*2*p1*t_I(m);
    end
    for k=1:(length(t_I)-1)
        Func_I(k)=(f_I(k)+f_I(k+1))/2;
    end
    alpha_I=(1/(c(i)*1000))*sum(Func_I');
    if alpha_I-(1-cl)==0 | (bb-aa)/2<tol
        c_naiman=c(i);
        alpha=alpha_I;break
    end
    i=i+1;
    if alpha_aa*(alpha_I-(1-cl))>0
```

```
        aa=c(i-1);
        alpha_aa=alpha_I;
    else
        bb=c(i-1);
    end
end
return
```

A.1.5 Obtaining c, using the approximate method for polynomial regression

```
function c_app=approxi_1d_cc(w,cl,p1,a,b);

%%Output
%c_app: the critical value of the approximate method
%%Input
%w,cl,p1,a,b: the same as before

n=length(w);
nu=n-p1;
for j=1:p1
    X(:,j)=w'.^(j-1);
end
q=X*inv(X'*X);
l=fliplr(q);
for j=1:n
    sql(j,:)=conv(l(j,:),l(j,:));
end
sqmoll=sum(sql);

x=a:(b-a)/1000:b;
for k=1:length(x)
```

```
    for j=1:(p1-1)
        xdl(j)=x(k)^(p1-1-j);
    end
    for j=1:((p1-1)*2+1)
        xsqmoll(j)=x(k)^((p1-1)*2+1-j);
    end
    for j=1:p1
        xl(j)=x(k)^(p1-j);
    end
    for j=1:(p1-1)*2
        xdsqmoll(j)=x(k)^((p1-1)*2-j);
    end
    for j=1:n
        dT(j)=(polyder(l(j,:))*xdl'*sqrt(sqmoll*xsqmoll')-
        l(j,:)*xl'*(1/(2*sqrt(sqmoll*xsqmoll')))*
        (polyder(sqmoll)*xdsqmoll'))/(sqmoll*xsqmoll');
    end
    moldT(k)=norm(dT);
end
for m=1:(length(x)-1)
    T(m)=(moldT(m)+moldT(m+1))/2;
end
k0=((b-a)/1000)*sum(T');

tol=0.0001;
N0=10000;
aa=1;bb=10;
alpha_aa=(k0/pi)*(1+aa^2/nu)^(-nu/2)+2*(1-tcdf(aa,nu));
i=1;
while i<=N0
    c(i)=aa+(bb-aa)/2;
```

```
    alpha_I=(k0/pi)*(1+c(i)^2/nu)^(-nu/2)+2*(1-tcdf(c(i),nu));
    if alpha_I-(1-cl)==0 | (bb-aa)/2<tol
        c_app=c(i);
        alpha=alpha_I;break
    end
    i=i+1;
    if alpha_aa*(alpha_I-(1-cl))>0
        aa=c(i-1);
        alpha_aa=alpha_I;
    else
        bb=c(i-1);
    end
end
return
```

A.1.6 Obtaining c, using the simulation-based method for polynomial regression

For quadratic regression

```
function [c_simu,se]=simu_quadratic_c(w,a,b,cl,nsim);

%%output
%c_simu: the critical value of the simulation-based method
%se: the standard error of c_simu
%%input
%w,a,b,cl: the same as before
%nsim: the number of simulations

n=length(w);
for m=1:3
    X(:,m)=w'.^(m-1);
```

```
end
q=inv(X'*X);
P=sqrtm(q);nu=n-3;

%Generate (beta^-beta)/sigma
V=P*randn(3,nsim);

%Generate sigma^/sigma
S=sqrt(sum(randn(nu,nsim).^2)./nu);

%Compute T
for m=1:nsim
    U=V(:,m)*V(:,m)';
    a1=U(2,1);
    a2=2*U(3,1)+U(2,2);
    a3=2*U(3,2)+U(2,3);
    a4=2*U(3,3);
    poly1=[a4 a3 a2 a1];

    b1=q(1,1);
    b2=q(2,1)+q(1,2);
    b3=q(3,1)+q(2,2)+q(1,3);
    b4=q(3,2)+q(2,3);
    b5=q(3,3);
    poly2=[b5 b4 b3 b2 b1];

    c1=U(1,1);
    c2=U(2,1)+U(1,2);
    c3=U(3,1)+U(2,2)+U(1,3);
    c4=U(3,2)+U(2,3);
    c5=U(3,3);
```

```
    poly3=[c5 c4 c3 c2 c1];

    d1=q(2,1);
    d2=2*q(3,1)+q(2,2);
    d3=2*q(3,2)+q(2,3);
    d4=2*q(3,3);
    poly4=[d4 d3 d2 d1];

    g=conv(poly1,poly2)-conv(poly3,poly4);
    y=roots(g);
    for j=1:3
        A(j)=a^(j-1);
        B(j)=b^(j-1);
    end
    ha=abs(A*V(:,m))/sqrt(A*q*A');
    hb=abs(B*V(:,m))/sqrt(B*q*B');
    for j=1:(4*3-6)
        for k=1:3
            Yj(k)=y(j)^(k-1);
        end
        if y(j)>a & y(j)<b
            h(j)=abs(Yj*V(:,m))/sqrt(Yj*q*Yj');
        else
            h(j)=0;
        end
    end
    H=[real(h) ha hb];
    H_max=max(H);
    Q(m)=H_max;
end
T=Q./S;
```

```
%Compute the quantile of the simulated values
T=sort(T);
r=nsim*cl;
c_simu=T(r);

%Compute the standard error
d=0.01;
K=(c_simu-T)/d;
g=sum((1/(d*sqrt(2*pi)))*exp(-0.5*(K.^2)))/length(T);
se=sqrt((cl*(1-cl))/(g^2*length(T)));
return
```

For cubic regression

```
function [c_simu,se]=simu_cubic_c(w,a,b,cl,nsim);

%All outputs and inputs are the same as before

n=length(w);
for m=1:4
    X(:,m)=w'.^(m-1);
end
q=inv(X'*X);
P=sqrtm(q);
nu=n-4;

%Generate (beta^-beta)/sigma
V=P*randn(4,nsim);

%Generate sigma^/sigma
```

```
S=sqrt(sum(randn(nu,nsim).^2)./nu);

%compute T
for m=1:nsim
    U=V(:,m)*V(:,m)';
    a1=U(2,1);
    a2=2*U(3,1)+U(2,2);
    a3=3*U(4,1)+2*U(3,2)+U(2,3);
    a4=3*U(4,2)+2*U(3,3)+U(2,4);
    a5=3*U(4,3)+2*U(3,4);
    a6=3*U(4,4);
    poly1=[a6 a5 a4 a3 a2 a1];

    b1=q(1,1);
    b2=q(2,1)+q(1,2);
    b3=q(3,1)+q(2,2)+q(1,3);
    b4=q(4,1)+q(3,2)+q(2,3)+q(1,4);
    b5=q(4,2)+q(3,3)+q(2,4);
    b6=q(4,3)+q(3,4);
    b7=q(4,4);
    poly2=[b7 b6 b5 b4 b3 b2 b1];

    c1=U(1,1);
    c2=U(2,1)+U(1,2);
    c3=U(3,1)+U(2,2)+U(1,3);
    c4=U(4,1)+U(3,2)+U(2,3)+U(1,4);
    c5=U(4,2)+U(3,3)+U(2,4);
    c6=U(4,3)+U(3,4);
    c7=U(4,4);
    poly3=[c7 c6 c5 c4 c3 c2 c1];
```

```
    d1=q(2,1);
    d2=2*q(3,1)+q(2,2);
    d3=3*q(4,1)+2*q(3,2)+q(2,3);
    d4=3*q(4,2)+2*q(3,3)+q(2,4);
    d5=3*q(4,3)+2*q(3,4);
    d6=3*q(4,4);
    poly4=[d6 d5 d4 d3 d2 d1];

    g=conv(poly1,poly2)-conv(poly3,poly4);
    y=roots(g);
    for j=1:4
        A(j)=a^(j-1);
        B(j)=b^(j-1);
    end
    ha=abs(A*V(:,m))/sqrt(A*q*A');
    hb=abs(B*V(:,m))/sqrt(B*q*B');
    for j=1:(4*4-6)
        for k=1:4
            Yj(k)=y(j)^(k-1);
        end
        if y(j)>a & y(j)<b
            h(j)=abs(Yj*V(:,m))/sqrt(Yj*q*Yj');
        else
            h(j)=0;
        end
    end
    H=[real(h) ha hb];
    H_max=max(H);
    Q(m)=H_max;
end
T=Q./S;
```

```
%Compute the quantile of the simulated values
T=sort(T);
r=nsim*cl;
c_simu=T(r);

%Compute the standard error
d=0.01;
K=(c_simm-T)/d;
g=sum((1/(d*sqrt(2*pi)))*exp(-0.5*(K.^2)))/length(T);
se=sqrt((cl*(1-cl))/(g^2*length(T)));
return
```

For 4th order polynomial regression

```
function [c_simu,se]=simu_poly_c(w,a,b,cl,nsim);

%All outputs and inputs are the same as before

n=length(w);
for m=1:5
    X(:,m)=w'.^(m-1);
end
q=inv(X'*X);
P=sqrtm(q);
nu=n-5;

%Generate (beta^-beta)/sigma
V=P*randn(5,nsim);

%Generate sigma^/sigma
```

```
S=sqrt(sum(randn(nu,nsim).^2)./nu);

%compute T
for m=1:nsim
    U=V(:,m)*V(:,m)';
    a1=U(2,1);
    a2=2*U(3,1)+U(2,2);
    a3=3*U(4,1)+2*U(3,2)+U(2,3);
    a4=4*U(5,1)+3*U(4,2)+2*U(3,3)+U(2,4);
    a5=4*U(5,2)+3*U(4,3)+2*U(3,4)+U(2,5);
    a6=4*U(5,3)+3*U(4,4)+2*U(3,5);
    a7=4*U(5,4)+3*U(4,5);
    a8=4*U(5,5);
    poly1=[a8 a7 a6 a5 a4 a3 a2 a1];

    b1=q(1,1);
    b2=q(2,1)+q(1,2);
    b3=q(3,1)+q(2,2)+q(1,3);
    b4=q(4,1)+q(3,2)+q(2,3)+q(1,4);
    b5=q(5,1)+q(4,2)+q(3,3)+q(2,4)+q(1,5);
    b6=q(5,2)+q(4,3)+q(3,4)+q(2,5);
    b7=q(5,3)+q(4,4)+q(3,5);
    b8=q(5,4)+q(4,5);
    b9=q(5,5);
    poly2=[b9 b8 b7 b6 b5 b4 b3 b2 b1];

    c1=U(1,1);
    c2=U(2,1)+U(1,2);
    c3=U(3,1)+U(2,2)+U(1,3);
    c4=U(4,1)+U(3,2)+U(2,3)+U(1,4);
    c5=U(5,1)+U(4,2)+U(3,3)+U(2,4)+U(1,5);
```

```
c6=U(5,2)+U(4,3)+U(3,4)+U(2,5);
c7=U(5,3)+U(4,4)+U(3,5);
c8=U(5,4)+U(4,5);
c9=U(5,5);
poly3=[c9 c8 c7 c6 c5 c4 c3 c2 c1];

d1=q(2,1);
d2=2*q(3,1)+q(2,2);
d3=3*q(4,1)+2*q(3,2)+q(2,3);
d4=4*q(5,1)+3*q(4,2)+2*q(3,3)+q(2,4);
d5=4*q(5,2)+3*q(4,3)+2*q(3,4)+q(2,5);
d6=4*q(5,3)+3*q(4,4)+2*q(3,5);
d7=4*q(5,4)+3*q(4,5);
d8=4*q(5,5);
poly4=[d8 d7 d6 d5 d4 d3 d2 d1];

g=conv(poly1,poly2)-conv(poly3,poly4);
y=roots(g);
for j=1:5
    A(j)=a^(j-1);
    B(j)=b^(j-1);
end
ha=abs(A*V(:,m))/sqrt(A*q*A');
hb=abs(B*V(:,m))/sqrt(B*q*B');
for j=1:(4*5-6)
    for k=1:5
        Yj(k)=y(j)^(k-1);
    end
    if y(j)>a & y(j)<b
        h(j)=abs(Yj*V(:,m))/sqrt(Yj*q*Yj');
    else
```

```
            h(j)=0;
        end
    end
    H=[real(h) ha hb];
    H_max=max(H);
    Q(m)=H_max;
end
T=Q./S;

%Compute the quantile of the simulated values
T=sort(T);
r=nsim*cl;
c_simu=T(r);

%Compute the standard error
d=0.01;
K=(c_simu-T)/d;
g=sum((1/(d*sqrt(2*pi)))*exp(-0.5*(K.^2)))/length(T);
se=sqrt((cl*(1-cl))/(g^2*length(T)));
return
```

A.1.7 Obtaining c, using the approximate method for bivariate linear regression

```
function cc_app=approxi_2d_cc(X,a,b,c,d,cl)

%%Output
%cc_app: the critical value of the approximate method
%%Input
%X: the design matrix
%a,b: the lower and upper bounds of the restricted interval
%c,d: the lower and upper bounds of the restricted interval
```

```
%cl: the confidence level

n=length(X(:,1)');
nu=n-3;
l=X*inv(X'*X);
for j=1:n
    sql(j,:)=[l(j,1)^2 l(j,2)^2 l(j,3)^2 2*l(j,1)*l(j,2)
     2*l(j,1)*l(j,3) 2*l(j,2)*l(j,3)];
end
sqmoll=sum(sql);
dxl=l(:,2);
dyl=l(:,3);
dxsqmoll=[2*sqmoll(2) sqmoll(4) sqmoll(6)];
dysqmoll=[2*sqmoll(3) sqmoll(5) sqmoll(6)];
x=a:(b-a)/1000:b;
y=c:(d-c)/1000:d;

%Compute k0
for j=1:length(x)
    for k=1:length(y)
        xysqmoll=[1 x(j)^2 y(k)^2 x(j) y(k) x(j)*y(k)];
        xyl=[1 x(j) y(k)];
        xydxsqmoll=[x(j) 1 y(k)];
        xydysqmoll=[y(k) 1 x(j)];
        Tx=(dxl*sqrt(sqmoll*xysqmoll')-(l*xyl')*
        (1/(2*sqrt(sqmoll*xysqmoll')))*
        (dxsqmoll*xydxsqmoll'))/(sqmoll*xysqmoll');
        Ty=(dyl*sqrt(sqmoll*xysqmoll')-(l*xyl')*
        (1/(2*sqrt(sqmoll*xysqmoll')))*
        (dysqmoll*xydysqmoll'))/(sqmoll*xysqmoll');
        A=[Tx Ty];
```

```
        f(k)=sqrt(det(A'*A));
    end
    for m=1:(length(y)-1)
        g(m)=(f(m)+f(m+1))/2;
    end
    k_int(j)=(d-c)/1000*sum(g');
end
for ii=1:(length(x)-1)
    K_int(ii)=(k_int(ii)+k_int(ii+1))/2;
end
k0=((b-a)/1000)*sum(K_int');

%Compute s0
for k=1:length(y)
    aysqmoll=[1 a^2 y(k)^2 a y(k) a*y(k)];
    ayl=[1 a y(k)];
    aydxsqmoll=[a 1 y(k)];
    aydysqmoll=[y(k) 1 a];
    Ty=(dyl*sqrt(sqmoll*aysqmoll')-(l*ayl')*
    (1/(2*sqrt(sqmoll*aysqmoll')))*
    (dysqmoll*aydysqmoll'))/(sqmoll*aysqmoll');
    f(k)=sqrt(det(Ty'*Ty));
end
for m=1:(length(y)-1)
    g(m)=(f(m)+f(m+1))/2;
end
ka_int=((d-c)/1000)*sum(g');

for k=1:length(y)
    bysqmoll=[1 b^2 y(k)^2 b y(k) b*y(k)];
    byl=[1 b y(k)];
```

```
    bydxsqmoll=[b 1 y(k)];
    bydysqmoll=[y(k) 1 b];
    Ty=(dyl*sqrt(sqmoll*bysqmoll')-(l*byl')*
    (1/(2*sqrt(sqmoll*bysqmoll')))*
    (dysqmoll*bydysqmoll'))/(sqmoll*bysqmoll');
    f(k)=sqrt(det(Ty'*Ty));
end
for m=1:(length(y)-1)
    g(m)=(f(m)+f(m+1))/2;
end
kb_int=((d-c)/1000)*sum(g');

for j=1:length(x)
    xcsqmoll=[1 x(j)^2 c^2 x(j) c x(j)*c];
    xcl=[1 x(j) c];
    xcdxsqmoll=[x(j) 1 c];
    xcdysqmoll=[c 1 x(j)];
    Tx=(dxl*sqrt(sqmoll*xcsqmoll')-(l*xcl')*
    (1/(2*sqrt(sqmoll*xcsqmoll')))*
    (dxsqmoll*xcdxsqmoll'))/(sqmoll*xcsqmoll');
    f(j)=sqrt(det(Tx'*Tx));
end
for m=1:(length(x)-1)
    g(m)=(f(m)+f(m+1))/2;
end
kc_int=((b-a)/1000)*sum(g');

for j=1:length(x)
    xdsqmoll=[1 x(j)^2 d^2 x(j) d x(j)*d];
    xdl=[1 x(j) d];
    xddxsqmoll=[x(j) 1 d];
```

```
    xddysqmoll=[d 1 x(j)];
    Tx=(dxl*sqrt(sqmoll*xdsqmoll')-(l*xdl')*
    (1/(2*sqrt(sqmoll*xdsqmoli')))*
    (dxsqmoll*xddxsqmoll'))/(sqmoll*xdsqmoll');
    f(j)=sqrt(det(Tx'*Tx));
end
for m=1:(length(x)-1)
    g(m)=(f(m)+f(m+1))/2;
end
kd_int=((b-a)/1000)*sum(g');
s0=ka_int+kb_int+kc_int+kd_int;

tol=0.0001;
N0=10000;
aa=1;bb=10;
alpha_aa=(k0/pi^(3/2))*(gamma((nu+1)/2)/gamma(nu/2))*(aa/sqrt(nu))*
(1+aa^2/nu)^(-(nu+1)/2)+(s0/(2*pi))*(1+aa^2/nu)^(-nu/2)+2*(1-tcdf(aa,nu));
i=1;
while i<=N0
    c(i)=aa+(bb-aa)/2;
    alpha_I=(k0/pi^(3/2))*(gamma((nu+1)/2)/gamma(nu/2))*(c(i)/sqrt(nu))*
    (1+c(i)^2/nu)^(-(nu+1)/2)+(s0/(2*pi))*(1+c(i)^2/nu)^(-nu/2)+
    2*(1-tcdf(c(i),nu));
    if alpha_I-(1-cl)==0 | (bb-aa)/2<tol
        cc_app=c(i);
        alpha=alpha_I;break
    end
    i=i+1;
    if alpha_aa*(alpha_I-(1-cl))>0
        aa=c(i-1);
        alpha_aa=alpha_I;
```

```
    else
        bb=c(i-1);
    end
end
return
```

A.2 For computing the simulated coverage probability for logistic regression

A.2.1 Obtaining *scp*, using the WB method for simple logistic regression

```
function scp_wb=wb_scp(x,N,b,a1,a2,nsim)

%%Output
%scp_wb: the simulated coverage probability of WB method
%%Input
%x: the design points of the only predictor variable
%N: the vector of sub-sample sizes
%b: the vector of true regression coefficients
%a1: the lower bound of the restricted interval
%a2: the upper bound of the restricted interval
%cl: the confidence level
%nsim: the number of simulations

n=length(x);
for k=1:nsim
    for j=1:n
        p_i(j)=exp(b(1)+b(2)*x(j))./(1+exp(b(1)+b(2)*x(j)));
        z(j)=binornd(N(j),p_i(j),1,1);
        if z(j)==N(j) | z(j)==0
```

```
        z(j)=binornd(N(j),p_i(j),1,1);
    end
end
y=z./N;
diff=1;
b_es=[0;0];  %initial guess of b_es
while diff>0.0001
    b_old=b_es;
    p=exp(b_es(1)+b_es(2)*x)./(1+exp(b_es(1)+b_es(2)*x));
    for i=1:length(x)
        J1(i)=N(i)*p(i)*(1-p(i));
        J2(i)=J1(i)*x(i);
        J3(i)=J2(i)*x(i);
    end
    s=[sum(y-p);sum((y-p).*x)];
    J=[sum(J1) sum(J2);sum(J2) sum(J3)];
    b_es=b_old+J\s;
    diff=sum(abs(b_es-b_old));
end
f_inv=inv(J);
P=sqrtm(f_inv);
vector_a=(P*[1;a1])';
vector_b=(P*[1;a2])';
theta_ast=acos((vector_a*vector_b')/(norm(vector_a)*
norm(vector_b)))./2;

tol=0.0001;
N0=10000;
aa=1;bb=10;
f=@(w)chi2cdf(aa.^2./(cos(w).^2),2);
g=quad(f,0,pi/2-theta_ast);
```

```
HA=(2/pi)*theta_ast*chi2cdf(aa^2,2)+(2/pi)*g;
i=1;
while i<=NO
    c(i)=aa+(bb-aa)/2;
    f1=@(y1)chi2cdf(c(i).^2./(cos(y1).^2),2);
    g1=quad(f1,0,pi/2-theta_ast);
    HI=(2/pi)*theta_ast*chi2cdf(c(i)^2,2)+(2/pi)*g1;
    if HI-0.95==0 | (bb-aa)/2<tol
        cc_wb=c(i);break
    end
    i=i+1;
    if HA*(HI-0.95)<0
       aa=c(i-1);
       HA=HI;
    else
        bb=c(i-1);
    end
end
v1=P(:,1)+P(:,2).*a1;
v2=P(:,1)+P(:,2).*a2;
M=inv(P)'*(b'-b_es);
if (M>=v1 & M<=v2) | (-M>=v1 & -M<=v2)
    Q=norm(M);
else
    Q1=abs(v1'*M)/norm(v1);
    Q2=abs(v2'*M)/norm(v2);
    Q=max(Q1,Q2);
end
T=Q;
if T>cc_wb
    r(k)=0;
```

```
    else
        r(k)=1;
    end
end
r_sum=sum(r');
scp_wb=r_sum/nsim;
return
```

A.2.2 Obtaining *scp* of Type 4 band for simple logistic regression, using **parfit** on S-plus

```
library(locfit,first=T)

scpT4<-function(x,N,b,a1,a2,nsim)
{
  for(i in 1:nsim)
  {
    pr<-exp(b[1]+b[2]*x)/(1+exp(b[1]+b[2]*x))
    z<-c(0,0,0,0,0)
    y<-c(0,0,0,0,0)
    J1<-c(0,0,0,0,0)
    J2<-c(0,0,0,0,0)
    J3<-c(0,0,0,0,0)
    v<-c(0,0,0)
    v1<-c(0,0,0)
    cc<-c(0,0,0)
    for(i in 1:length(x))
    {
      z[i]<-rbinom(1,N[i],pr[i])
    }
    y<-z/N
    bb<-glm(y~x,family=binomial)
```

```
bes<-unlist(bb[1],use.names=F)
p<-exp(bes[1]+bes[2]*x)/(1+exp(bes[1]+bes[2]*x))
for(i in 1:length(x))
{
  J1[i]<-N[i]*p[i]*(1-p[i])
  J2[i]<-J1[i]*x[i]
  J3[i]<-J2[i]*x[i]
}
J0<-c(sum(J1),sum(J2),sum(J2),sum(J3))
J<-matrix(J0,nrow=2,byrow=T)
finv<-solve(J)
t<-data.frame(x,z,N)
fit<-scb(z~x,type=4,w=N,data=t,deg=1,family="binomial",
kern="parm",xlim=c(a1,a2))
xp<-unlist(fit[1],use.names=F)
ll<-unlist(fit[4],use.names=F)
ul<-unlist(fit[5],use.names=F)
for(i in c(1,10,20))
{
  cc[i]<-(ul[i]-ll[i])/(2*sqrt(c(1,xp[i])%*%finv%*%
  matrix(c(1,xp[i]),nrow=2)))
}
ccapp<-(cc[1]+cc[10]+cc[20])/3
R<-seq(0,by=0,length=nsim)
q<-0
while(q<=20)
{
  u<-a1+q*(a2-a1)/20
  G1<-c(1,u)%*%matrix(b-bes,nrow=2)
  G2<-sqrt(c(1,u)%*%finv%*%matrix(c(1,u),nrow=2))
  H<-abs(G1)/G2
```

```
      if(H>ccapp)
      {
        R[i]<-0
        break
      }
      R[i]<-1
      q<-q+1
    }
  }
  Rsum<-sum(R)
  scpapp<-Rsum/nsim
  return(scpapp)
}
```

A.2.3 Obtaining *scp*, using the simulation-based method for bivariate logistic regression

```
function scp_simu=simu_scp(x1,x2,N,b,a1,a2,a3,a4,cl,nsim1,nsim2)

%%Output
%scp_simu: the simulated coverage probability of the confidence
%          band constructed based on the simulation method
%%Input
%x1: the design points of the first predictor variable
%x2: the design points of the second predictor variable
%N: the vector of sub-sample sizes
%b: the vector of true regression coefficients
%a1: the lower bound of the first restricted interval
%a2: the upper bound of the first restricted interval
%a3: the lower bound of the second restricted interval
%a4: the upper bound of the second restricted interval
%cl: the confidence level
```

```
%nsim1: the number of simulations for computing critical value
%nsim2: the number of simulations for computing coverage probability

for k=1:nsim2
    p_i=exp(b(1)+b(2)*x1+b(3)*x2)./(1+exp(b(1)+b(2)*x1+b(3)*x2));
    for j=1:length(x1)
        z(j)=binornd(N(j),p_i(j),1,1);
        if z(j)==0 | z(j)==N(j)
           z(j)=binornd(N(j),p_i(j),1,1);
        end
    end
    y=z./N;
    diff=1;
    b_es=[0;0;0];  %initial guess of b_es
    while diff>0.0001
        b_old=b_es;
        pr=exp(b_es(1)+b_es(2)*x1+b_es(3)*x2)./
        (1+exp(b_es(1)+b_es(2)*x1+b_es(3)*x2));
        for i=1:length(x1)
            J1(i)=N(i)*pr(i)*(1-pr(i));
            J2(i)=J1(i)*x1(i);
            J3(i)=J2(i)*x1(i);
            J4(i)=J1(i)*x2(i);
            J5(i)=J2(i)*x2(i);
            J6(i)=J4(i)*x2(i);
        end
        s=[sum(y-pr);sum((y-pr).*x1);sum((y-pr).*x2)];
        J=[sum(J1) sum(J2) sum(J4);sum(J2) sum(J3) sum(J5);
        sum(J4) sum(J5) sum(J6)];
        b_es=b_old+J\s;
        diff=sum(abs(b_es-b_old));
```

```
end
f_inv=inv(J);
P=sqrtm(f_inv);
v1=P(:,1)+P(:,2).*a1+P(:,3).*a3;
v2=P(:,1)+P(:,2).*a1+P(:,3).*a4;
v3=P(:,1)+P(:,2).*a2+P(:,3).*a3;
v4=P(:,1)+P(:,2).*a2+P(:,3).*a4;
for i=1:nsim1
    M=randn(3,1);
    if (M>=v1 & M<=v4) | (-M>=v1 & -M<=v4)
        Q=norm(M);
    else
        B1=[v1 v2];
        B2=[v1 v3];
        B3=[v2 v4];
        B4=[v3 v4];
        [Q1,R1]=qr(B1,0);
        [Q2,R2]=qr(B2,0);
        [Q3,R3]=qr(B3,0);
        [Q4,R4]=qr(B4,0);
        D1=dot(Q1(:,1),M)*Q1(:,1)+dot(Q1(:,2),M)*Q1(:,2);
        D2=dot(Q2(:,1),M)*Q2(:,1)+dot(Q2(:,2),M)*Q2(:,2);
        D3=dot(Q3(:,1),M)*Q3(:,1)+dot(Q3(:,2),M)*Q3(:,2);
        D4=dot(Q4(:,1),M)*Q4(:,1)+dot(Q4(:,2),M)*Q4(:,2);
        if (D1>=v1 & D1<=v2) | (-D1>=v1 & -D1<=v2)
            Q11=abs(D1'*M)/norm(D1);
        else
            Q11=max(abs(v1'*M)/norm(v1),abs(v2'*M)/norm(v2));
        end
        if (D2>=v1 & D2<=v3) | (-D2>=v1 & -D2<=v3)
            Q12=abs(D2'*M)/norm(D2);
```

```
        else
            Q12=max(abs(v1'*M)/norm(v1),abs(v3'*M)/norm(v3));
        end
        if (D3>=v2 & D3<=v4) | (-D3>=v2 & -D3<=v4)
            Q13=abs(D3'*M)/norm(D3);
        else
            Q13=max(abs(v2'*M)/norm(v2),abs(v4'*M)/norm(v4));
        end
        if (D4>=v3 & D4<=v4) | (-D4>=v3 & -D4<=v4)
            Q14=abs(D4'*M)/norm(D4);
        else
            Q14=max(abs(v3'*M)/norm(v3),abs(v4'*M)/norm(v4));
        end
        Qarray=[Q11 Q12 Q13 Q14];
        Q=max(Qarray);
    end
    T(i)=Q;
end
T=sort(T);
r=nsim1*cl;
cc_simu=T(r);

MM=inv(P)'*(b'-b_es);
if (MM>=v1 & MM<=v4) | (-MM>=v1 & -MM<=v4)
    QQ=norm(MM);
else
    DD1=dot(Q1(:,1),MM)*Q1(:,1)+dot(Q1(:,2),MM)*Q1(:,2);
    DD2=dot(Q2(:,1),MM)*Q2(:,1)+dot(Q2(:,2),MM)*Q2(:,2);
    DD3=dot(Q3(:,1),MM)*Q3(:,1)+dot(Q3(:,2),MM)*Q3(:,2);
    DD4=dot(Q4(:,1),MM)*Q4(:,1)+dot(Q4(:,2),MM)*Q4(:,2);
    if (DD1>=v1 & DD1<=v2) | (-DD1>=v1 & -DD1<=v2)
```

```
            QQ11=abs(DD1'*MM)/norm(DD1);
        else
            QQ11=max(abs(v1'*MM)/norm(v1),abs(v2'*MM)/norm(v2));
        end
        if (DD2>=v1 & DD2<=v3) | (-DD2>=v1 & -DD2<=v3)
            QQ12=abs(DD2'*MM)/norm(DD2);
        else
            QQ12=max(abs(v1'*MM)/norm(v1),abs(v3'*MM)/norm(v3));
        end
        if (DD3>=v2 & DD3<=v4) | (-DD3>=v2 & -DD3<=v4)
            QQ13=abs(DD3'*MM)/norm(DD3);
        else
            QQ13=max(abs(v2'*MM)/norm(v2),abs(v4'*MM)/norm(v4));
        end
        if (DD4>=v3 & DD4<=v4) | (-DD4>=v3 & -DD4<=v4)
            QQ14=abs(DD4'*MM)/norm(DD4);
        else
            QQ14=max(abs(v3'*MM)/norm(v3),abs(v4'*MM)/norm(v4));
        end
        QQarray=[QQ11 QQ12 QQ13 QQ14];
        QQ=max(QQarray);
    end
    TT=QQ;
    if TT>cc_simu
        rr(k)=0;
    else
        rr(k)=1;
    end
end
rr_sum=sum(rr');
scp_simu=rr_sum/nsim2; return
```

References

1. Al-Saidy, O.M., Piegorsch, W.W., West, R.W. et al. (2002). "Confidence bands for low-dose risk estimation with quantal response data," *Biometrics*, **59**, 1056-1062.

2. Agresti, A., and Coull, B.A. (1998). "Approximate is better than 'exact' for interval estimation of binomial proportions," *American Statistician*, **52**, 119-126.

3. Blyth, C.R., and Still, H.A. (1983). "Binomial confidence intervals," *Journal of the American Statistical Association*, **78**, 108-116.

4. Bohrer, R. (1967). "On sharpening Scheffé bounds," *Journal of the Royal Statistical Society (B)*, **29**, 110-114.

5. Bohrer, R. and Francis, G.K. (1972). "Sharp one-sided confidence bands for linear regression over intervals," *Biometrika*, **59**, 99-107.

6. Bohrer, R. (1973). "A multivariate t probability integral," *Biometrika*, **60**, 647-654.

7. Bowden, D.C. (1970). "Simultaneous confidence bands for linear regression models," *Journal of the American Statistical Association*, **65**, 413-421.

8. Brand, R.J., Pinnock, D.E., and Jackson, K.L. (1973). "Large sample confidence bands for the logistic response curve and its inverse," *American Statistician*, **27**, 157-160.

9. Brown, L.D., Cai, T.T., and DasGupta, A. (2000). "Interval estimation in exponential families," Technical report, available at www-stat.wharton.upenn.edu/~tcai/.

10. Brown, L.D., Cai, T.T., and DasGupta, A. (2001). "Interval estimation for a binomial proportion," *Statistical Science*, **16**, 101-133.

11. Brown, L.D., Cai, T.T., and DasGupta, A. (2002). "Confidence intervals for a binomial proportion and asymptotic expansions," *Annals of Statistics*, **30**, 160-201.

12. Casella, G. and Strawderman, W.E. (1980). "Confidence bands for linear-regression with restricted predictor variables," *Journal of the American Statistical Association*, **75**, 862-868.

13. Chen, H. (1990). "The accuracy of approximate intervals for a binomial parameter," *Journal of the American Statistical Association*, **85**, 514-518.

14. Clopper, C.J., and Pearson, E.S. (1934). "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, **26**, 404-413.

15. Collett, D. (2002). "*Modelling Binary Data*," 2nd ed, Chapman & Hall/CRC.

16. Dobson, A.J. (2001). "*An Introduction to Generalized Linear Models*," 2nd ed, Chapman & Hall/CRC.

17. Gafarian, A.V. (1964). "Confidence bands in straight line regression," *Journal of the American Statistical Association*, **59**, 182-213.

18. Ghosh, B.K. (1979). "A comparison of some approximate confidence intervals for the binomial parameter," *Journal of the American Statistical Association*, **74**, 894-900.

19. Graybill, F.A. and Bowden, D.C. (1967). "Linear segment confidence bands for simple linear regression models," *Journal of the American Statistical Association*, **62**, 403-408.

20. Halperin, M. and Gurian, J. (1968). "Confidence bands in linear regression with constraints on independent variables," *Journal of the American Statistical Association*, **63**, 1020-1027.

21. Halperin, M., Rastogi, S.C., Ho, I., and Yang, Y.Y. (1967). "Shorter confidence bands in linear regression," *Journal of the American Statistical Association*, **62**, 1050-1067.

22. Hauck, W.W. (1983). "A note on confidence bands for the logistic response curve," *American Statistician*, **37**, 158-160.

23. Hayter, A.J., Liu, W., and Wynn, H.P. (2005). "Easy-to-construct confidence bands for comparing two simple linear regression lines," *Manuscript*, Georgia Tech, USA.

24. Hsu, J.C. (1996). "*Multiple Comparisons - Theory and Methods*," Chapman & Hall.

25. Johansen, S. and Johnstone, I. (1990). "Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis," *Annals of Statistics*, **18**, 652-684.

26. Knafl, G., Sacks, J., and Ylvisaker, D. (1985). "Confidence bands for regression-functions," *Journal of the American Statistical Association*, **80**, 683-691.

27. Kosorok, M.R. and Qu, R. (1999). "Exact simultaneous confidence bands for a collection of univariate polynomials in regression analysis," *Statistics in Medicine*, **18**, 613-620.

28. Liu, W., Jamshidian, M., Zhang, Y., and Bretz, F. (2004). "Constant width simultaneous confidence bands in multiple linear regression

with predictor variables constrained in intervals," *Journal of Statistical Computation & Simulation*, **00**, 1-12.

29. Liu, W., Jamshidian, M., Zhang, Y., and Donnelly, J. (2005). "Exact simultaneous confidence bands in multiple linear regression with predictor variables constrained in intervals," *Journal of Computational and Graphical Statistics*, **14(2)**, 459-484.

30. Liu, W., Wynn, H.P., and Hayter, A.J. (2005). "Statistical inferences for polynomial regression models."

31. Loader, C. (2004). "The Volume-of-Tube formula: Computational methods and statistical applications."

32. Montgomery, D.C., Peck, E.A., and Vining, G.G. (2006). "*Introduction to Linear Regression Analysis*," 4th ed, Wiley.

33. Naiman, D.Q. (1983). "Comparing Scheffé-type to constant-width confidence bounds in regression," *Journal of the American Statistical Association*, **78**, 906-912.

34. Naiman, D.Q. (1986). "Conservative confidence bands in curvilinear regression," *Annals of Statistics*, **14**, 896-906.

35. Naiman, D.Q. (1987). "Simultaneous confidence-bounds in multiple-regression using predictor variable constraints," *Journal of the American Statistical Association*, **82**, 214-219.

36. Naiman, D.Q. (1990). "Volumes of tubular neighborhoods of spherical polyhedra and statistical inference," *Annals of Statistics*, **18**, 685-716.

37. Piegorsch, W.W. and Casella, G. (1988). "Confidence bands for logistic regression with restricted predictor variables," *Biometrics*, **44**, 739-750.

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

39. Seppanen, E. and Uusipaikka, E. (1992). "Confidence bands for linear-regression over restricted regions," *Scandinavia Journal of Statistics*, **19**, 73-81.

40. Snee, R.D. (1977). "Validation of regression models: methods and examples," *Technometrics*, **19**, 415-428.

41. Sun, J.Y. and Loader, C.R. (1994). "Simultaneous confidence bands for linear regression and smoothing," *Annals of Statistics*, **22**, 1328-1346.

42. Sun, J.Y., Loader, C.R., and McCormick, W.P. (2000). "Confidence bands in generalized linear models," *Annals of Statistics*, **28**, 429-460.

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

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

45. Weisberg, S. (2005). "*Applied Linear Regression*," 3rd ed, Wiley.

46. Wilson, E.B. (1927). "Probable inference, the law of succession, and statistical inference," *Journal of the American Statistical Association*, **22**, 209-212.

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

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

49. Wynn, H.P. (1975). "Integrals for one-sided confidence bounds: A general result," *Biometrika*, **62**, 393-396.

50. Wynn, H.P. (1984). "An exact confidence band for one-dimensional polynomial regression," *Biometrika*, **71**, 375-379.

51. Yeh, B. (1996). "Bootstrap percentile confidence bands based on the concept of curve depth," *Communications in Statistics - Simulation and computation*, **25**, 905-922.

52. Zheng, X.D., and Loh, W.Y. (1995). "Bootstrapping binomial confidence-intervals," *Journal of Statistical Planning and Inference*, **43**, 355-380.

53. Zheng, X.D. (1998). "Better saddlepoint confidence intervals via bootstrap calibration," *American Mathematical Society*, **126**, 3669-3679.